Nonclassical Measurement Error in a Nonlinear (Duration) Model

Daniel Gutknecht

No 961

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Nonclassical Measurement Error in a Nonlinear (Duration) Model

Daniel Gutknecht

July 1, 2011

Abstract

In this paper, we study nonclassical measurement error in the continuous dependent variable of a semiparametric transformation model. The latter is a popular choice in practice nesting various nonlinear duration and censored regression models. The main complication arises because we allow the (additive) measurement error to be correlated with a (continuous) component of the regressors as well as with the true, unobserved dependent variable itself. This problem has not yet been studied in the literature, but we argue that it is relevant for various empirical setups with mismeasured, continuous survey data like earnings or durations. We develop a framework to identify and consistently estimate (up to scale) the parameter vector of the transformation model. Our estimator links a two-step control function approach of Imbens and Newey (2009) with a rank estimator similar to Khan (2001) and is shown to have desirable asymptotic properties. We prove that 'm out of n' bootstrap can be used to obtain a consistent approximation of the asymptotic variance and study the estimator's finite sample performance in a Monte Carlo Simulation. To illustrate the empirical usefulness of our procedure, we estimate an earnings equation model using annual data from the Health and Retirement Study (HRS). We find some evidence for a bias in the coefficients of years of education and age, emphasizing once again the importance to adjust for potential measurement error bias in empirical work.

JEL Classification: C14, C34,

Key-Words: Nonclassical Measurement Error, Dependent Variable, Control Function, Rank Estimator

Acknowledgements: I would like to thank my supervisors Valentina Corradi and Wiji Arulampalam for their encouraging support and helpful comments. I would also like to express my gratitude to S. Khan, Y. Shin, and E. Tamer for providing me with their GAUSS routines from the paper *Heteroscedastic Transformation Models with Covariate Dependent Censoring* (JBES 2011, 29(1), p.40-48). Comments received from seminar participants at the University of Warwick and the Econometric Study Group Meeting in Bristol (2009) are gratefully acknowledged.

University of Warwick, Department of Economics, Coventry CV4 7AL, UK, email: d.gutknecht@warwick.ac.uk

1 Introduction

The paper considers identification and estimation of the parameter vector of the monotone transformation model (Han, 1987) when the continuous dependent variable is subject to nonclassical measurement error, where 'nonclassical' refers to a potential correlation of the measurement error with the true, unobserved dependent variable itself and a (continuous) component of the regressor vector. This setup is of interest from an empirical perspective as survey data is commonly subject to measurement error (Bound, Brown, and Mathiowetz, 2001). In particular for earnings and duration data, which is often analysed using models nested within the monotone transformation model, evidence suggests that nonclassical measurement error is the rule rather than the exception: Bricker and Engelhardt (2007) for instance study measurement error in matched earnings data of older workers in the Health and Retirement Study (HRS). Their findings suggest a strong negative ('mean-reverting') relationship between the extent of measurement error and true earnings. In addition, this measurement error is found to rise with reported (years of) education. Cristia and Schwabish (2007) confirm both results using the Survey of Income and Program Participation (SIPP) Panel matched to administrative records.[1] In the duration context, Jaeckle (2008) reveals a similar pattern for benefit recipient histories in the British Household Panel Survey (BHPS), where individuals with lower educational attainment tend to over-report (i.e. to report durations that exceeded the ones they actually experienced) more excessively and under-reporting of the benefit duration generally increases with spell length. Since both, durations and earnings, are typically used as 'left-hand side' variables in censored regression or duration models, both examples can be accomodated by our framework.[2]

The main contribution of this paper is to provide the researcher with a tool to deal with nonclassical (as defined above) measurement error in continuous survey data such as earnings or durations if the model of interest is the monotone transformation model (or any other model nested therein). To the best of our knowledge, such a tool does not yet exist. While correlation between the unobserved dependent variable and measurement error seems to be the more common finding in practice (see Torelli and Trivellato, 1989; Bound and Krueger, 1991), the main theoretical complication in the identification and estimation process of the parameter vector actually arises from the correlation of the measurement error and the continuous component of the regressors. We solve this 'endogeneity problem' by proposing a three-step identification and estimation procedure: first, we employ a two-step control function approach (see Imbens and Newey, 2009; Hahn, Hu, and Ridder, 2008) to estimate the conditional mean of our (mismeasured) dependent variable conditional on all covariates and the estimated control function.

[1] In addition, the study also provides evidence for a correlation of measurement error with other demographic variables such as gender, age, or marital status.

[2] Notice that in order to apply the framework of this paper, education needs to be modelled as a continuous variable ('years of schooling').

Subsequently integrating over the support of the control function reduces the measurement error to a numerical constant. In a third step, we then use a rank-type argument comparing pairs of observations to eliminate this numerical constant. Since we employ a control function method in the first place, our procedure requires the existence of a suitable instrument vector. We argue that for the examples given before, instrumental variables typically suggested by the empirical literature for Mincer-type earnings equations such as parental education, minimum school-leaving age, or (same sex) sibling's educational qualification should also be applicable in this context as they are likely to be correlated with the observed schooling level of the individual, but unlikely to affect the individual's actual response to the survey question (see section 2.1).[3]

From a technical point of view, the main innovation of the paper is to combine a nonparametric mean estimator with a rank estimation procedure and to derive its asymptotic properties.[4] Since duration models are arguably one of the most relevant application field of the transformation model in practice, we extend the estimator to allow for random right censoring. The additional estimation step required to accomodate censoring and to obtain the mean function further complicate the asymptotic variance expression, which depends on first and second order derivatives of certain conditional expectations. Thus, in order to construct confidence intervals for our parameter estimates, we suggest the use of 'm out of n' bootstrap for these intervals and show its first order validity. Finally, to illustrate our methodology empirically, we examine annual earnings data from the HRS, which has been found to be subject to nonclassical measurement error (Bricker and Engelhardt, 2007). We estimate the reduced version of an earnings equation and find that our estimator differs substantially from other estimators obtained for comparison purposes. Together with evidence for a mean-reverting non-classical measurement error in annual earnings in the HRS (see Bricker and Engelhardt, 2007), this underlines the need to adjust for measurement error bias when examining the determinants of annual labour income of older workers in the HRS as estimates appear to be strongly affected.

Our paper complements the existing literature on nonlcassical measurement error, which has been rather limited regarding measurement error in the response variable of nonlinear models. Within the duration setup for instance, researchers have limited attention to either fully parametric duration models or classical forms of measurement error (e.g. Skinner and Humphreys, 1999; Augustin, 1999; Abrevaya and Hausman, 2004; Dumangane, 2007), both of which are problematic once the restrictive setup fails to hold. A notable exception is the paper by Abrevaya and Hausman (1999), who consider nonadditive, classical measurement error in the dependent variable. Relative to our approach, however, their setup cannot incorporate

[3]Notice that this argument is valid even when measurement error is actually not related to cognitive ability as in the earnings example in the introduction, where correlation with educational attainment is more likely to be due to its link with higher earnings.

[4]Concurrently to this work, Jochmans (2010) developed a two-step weighted rank estimator with nonparametric controls for the monotone transformation model.

a correlation of the measurement error with the true, unobserved dependent variable itself, which often appears to be the more relevant problem in practice. Abstracting from the duration context, Chen, Hong, and Tamer (2005) have considered various semiparametric models under nonclassical measurement error (in the dependent as well as the independent variable(s)) using auxiliary data. Matzkin (2007) examines a completely nonparametric framework, but her identification result hinges on the independence of the response error and other model (un-)observables. Hoderlein and Winter (2007) on the other hand use a structural approach to identify marginal effects of linear and nonlinear models under measurement error in either the dependent or the independent variable(s). While their methodology allows them to make detailed statements about the determinants and implications of such a measurement error, the validity of these claims clearly relies on the underlying model assumptions.

The paper is organised as follows: Section 2.1 outlines the identification strategy. Section 2.2 deals with the corresponding multi-step estimation procedure, its asymptotic distribution is derived in Section 2.3 and the validity of the bootstrapped confidence intervals is established in Section 2.4. Finally, Section 3 explores the finite sample properties in a small scale simulation study and Section 4 concludes with an empirical illustration on annual earnings data from the HRS Survey. All tables and proofs are postponed to the appendix.

2 Setup

2.1 Identification

The monotone transformation model (Han, 1987), which nests several duration and censored regression models, is given by:

$$Y_j^* = m(X_j'\beta_0 + \epsilon_j) \tag{1}$$

where Y_j^* is an unobserved, continuous scalar dependent variable, $X_j' = \{X_j^{(c)}, X_j^{(d)}\}'$ is a $(d_x \times 1)$-dimensional covariate vector with $X_j^{(c)}$ $\left(X_j^{(d)}\right)$ containing continuous (discrete) elements, and ϵ_j is a scalar unobservable (independent of X_j). $m(\cdot)$ is a strictly increasing transformation function giving the model its flexibility and name.[5] Within this setup, we incorporate an additively separable[6], nonclassical measurement error η_j (scalar):

$$Y_j = Y_j^* + \eta_j \tag{2}$$

[5]Notice that the flexible structure of $m(\cdot)$ only allows identification up to a location and size normalization (Sherman, 1993).

[6]Notice that if Y_j^* is a duration variable, which can only take on positive values, the expression in (2) can be interpreted as the log-transformation of $\widetilde{Y}_j = \widetilde{Y}_j^* \cdot \widetilde{\eta}_j$, where both $\widetilde{Y}_j^*$ and $\widetilde{\eta}_j$ have support $[0, \infty)$ and $\widetilde{Y}_j^*, \widetilde{\eta}_j > 0$ except for a set of measure zero. Thus, the assumption of additive separability is not as restrictive as it might appear at first sight and has in fact been adopted by several authors in the literature (e.g Chesher, Dumangane, and Smith, 2002).

'Nonclassical' here refers to a potential correlation of the measurement error with the true, underlying dependent variable and a (continuous) component of X_j. That is, letting "$\perp$" denote statistical independence of two random variables and "$\not\perp$" dependence, we have:

- $\epsilon_j \not\perp \eta_j$ and
- $X_{1j} \not\perp \eta_j$, where $X_{1j} \in X_j^{(c)}$.

The fact that we only allow for a correlation of the measurement error with one continuous component of X_j is due to the nonparametric regression at the first stage and certainly restrictive. We believe, however, that our setup is still covering a sufficiently large number of interesting applications (see introduction) to be of relevance. Combining (1) and (2) yields the observed equation:

$$Y_j = m(X_j'\beta_0 + \epsilon_j) + \eta_j \tag{3}$$

Our object is to identify β_0 from (3). We assume the existence of an instrument vector Z_j:

A1 there exists a $(d_z \times 1)$-dimensional, continuous vector Z_j (with $d_z \geq 1$) such that

$$X_{1j} = g(Z_j) + V_j \tag{4}$$

with $g(\cdot)$ a real-valued, continuous function, $\mathbb{E}[V_j|Z_j] = 0$, and

$$Z_j \perp \epsilon_j, \eta_j, V_j$$

.

Condition A1 is the "exclusion restriction" typically imposed in the control function literature. It specifies that the correlation between X_{1j} and η_j only runs through a function V_j, the so called control function. Moreover, notice that we require the instrumental variables to be continuous due to the use of kernel methods at the first-stage of our estimator.[7] Concerning the examples given in the introduction, instruments suggested in the context of (Mincer-type) earnings equations (Glewwe and Patrinos, 1999; Butcher and Case, 1994; Card, 2001; Ichino and Winter-Ebmer, 1999) should also be applicable here. However, in line with Hu and Schennach (2008), we stress that also a repeated measurement of Y_j^* could be understood as an instrument if it satisfied the independence assumption in A1 (i.e. if the second observation was independent of the measurement error η_j in Y_j conditional on the regressors X_j).[8] Finally, since our setup

[7] In practice, if some components of the instrument vector Z_j were discrete, nonparametric estimation of $g(\cdot)$ could proceed by splitting the sample according to the different values of that component and by estimating $g(\cdot)$ separately for these values.

[8] If one believes that an alternative measurement is independent of the original measurement error (possibly conditional on other variables), repeated measurements could be employed as suitable instruments. See Chalak

is entirely nonparametric, it is well known that identification condition A1 does not imply nor is it implied by the assumptions usually imposed in the nonlinear instrumental variable (NIV) literature.

The second condition we require for identification is a "large support condition", which ensures sufficient variation in V_j given X_{1j} (see Hahn, Hu, and Ridder (2008) for details).

A2 $\mathcal{W} = \mathcal{X} \times \mathcal{Z} \times \mathcal{V}$ is a compact, non-empty subset in the interior of the joint support of X, Z, and V. Assume that the joint density on $\mathcal{W}$ is everywhere continuous and bounded away from zero.

The third and fourth condition sufficient for identification of β_0 are:

A3 $\{X_j, Z_j, \epsilon_j, \eta_j\}_{j=1}^n$ is an i.i.d. sample, where Y_j and the endogenous component X_{1j} are generated according to (3) and (4), respectively.

A4 Let $\mu(x) := \int \mathbb{E}[Y_j|X_j = x, V_j = v] f_V(v)dv$ with $f_V(\cdot)$ the marginal density of V_j. For every $x \in \mathcal{X}$, $\mu(x)$ is strictly increasing in $x^{'}\beta_0$.

Condition A3 is a standard i.i.d. assumption, while A4 serves to make the restrictions imposed on $m(\cdot)$ and the index $x'\beta_0$ explicit.[9] That is, condition A4 follows from $Y_j = m(X_j'\beta_0 + \epsilon_j) + \eta_j$ with $m(\cdot)$ being strictly increasing in $X_j'\beta_0$ and the argument used in the proof of Lemma 1, which can be found in the appendix.

Given this setup, we obtain the following lemma, which ensures that the limit of the objective function introduced in the next section is uniquely maximized:

Lemma 1. *Under assumptions A1, A2, A3, and A4 and given (3) and (4), we have for every* $x, \widetilde{x} \in \mathcal{X}$*:*

$$\mu(x) > \mu(\widetilde{x}) \quad \textit{if} \quad x'\beta_0 > \widetilde{x}'\beta_0$$

The proof of this lemma proceeds in two steps: using A1 and A2 and an iterated expectation argument, one can show that for every $x \in \mathcal{X}$ (see Imbens and Newey (2009)):

$$\mu(x) = \mathbb{E}[m(x'\beta_0 + \epsilon_j)] + \mathbb{E}[\eta_j] \tag{5}$$

where the expectation is taken w.r.t. ϵ_j and η_j, respectively. Notice that $\mathbb{E}[\eta_j]$ is 'reduced' to a numerical constant and that $\mu(x)$, by A4, is strictly increasing in $x'\beta_0$ for all $x \in \mathcal{X}$. The latter

and White (2007) for a detailed discussion of identification under various instrument concepts.

[9] Notice that a further support condition similar to Cavanagh and Sherman (1998) will ensure that identification is not lost by restricting ourselves to the set $\mathcal{W}$.

motivates the use of a rank-type argument (see Cavanagh and Sherman, 1998), which together with the i.i.d. assumption A3 allows for identification of β_0. That is, by A3 we have for every $x \in \mathcal{X}$ and $i, j \in 1, \ldots, n$:

$$\mathbb{E}[m(x'\beta_0 + \epsilon_j)] + \mathbb{E}[\eta_j] = \mathbb{E}[m(x'\beta_0 + \epsilon_i)] + \mathbb{E}[\eta_i]$$

Thus, given x, an inequality will only arise for differing β-values.

2.2 Estimation

We suggest a three-step estimation procedure that follows from the previous result:

(i) Recover $\widehat{V}_j$ from a nonparametric first-stage regression of X_{1j} on Z_j.

(ii) Estimate $\mu(x, v) := \mathbb{E}[Y_j | X_j = x, V_j = v]$ nonparametrically using $Y_j, X_j, \widehat{V}_j$. Compute the average: $\widehat{\mu}(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}(x, \widehat{V}_i)$ for every $x \in \mathcal{X}$.

(iii) Use a modified version of the two-step rank estimator of Khan (2001) to recover β_0 (up to scale).

The last step requires the use of a modified version of Khan's (2001) rank estimator, which uses a conditional quantile function as transformation of the dependent variable. We replace this conditional quantile function and its estimator by the conditional mean $\mu(x)$ and $\widehat{\mu}(x)$, respectively. The replacement (together with the introduction of a control function and censoring) affects the asymptotic variance of our estimator, which will be different from the expression derived in Khan (2001). The estimated control functions $\widehat{V}_j$ stem from the regression equivalent of (4), that is:

$$\widehat{V}_j = X_{1j} - \widehat{g}(Z_j)$$

To estimate $g(\cdot)$, we use the Nadaraya-Watson estimator (for simplicity, assume that $d_z = 1$) with

$$\widehat{g}(Z_j) = \frac{\sum_{k=1}^{n} X_{1k} \mathbf{k}_h(Z_j - Z_k)}{\sum_{k=1}^{n} \mathbf{k}_h(Z_j - Z_k)}$$

where

$$\mathbf{k}_h(Z_j - Z_k) = \mathbf{k}\left(\frac{Z_j - Z_k}{h}\right)$$

and h is a deterministic sequence satisfying $h \longrightarrow 0$ as $n \longrightarrow \infty$, while $\mathbf{k}(\cdot)$ is a standard kernel function defined in B3 in Appendix A.1. Notice that $g(\cdot)$ could also be estimated by series estimators (splines, power series) or local linear smoothers, but the use of the Nadaraya Watson estimator will facilitate several proofs in the appendix. Moreover, we obtain a lim-

iting distribution that does not depend on the nonparametric first step estimators (a similar was obtained by Newey (1994) for smooth objective functions with a nonparametric plug-in estimate).

The conditional mean function $\mu(x)$ can be estimated using again the Nadaraya-Watson kernel estimator. Since we have a d_x-dimensional covariate vector X_j and a univariate $\widehat{V}_j$, we define the following $d = (d_x + 1)$ dimensional product kernel (for simplicity assume that: $h = h_1 = h_2 = ... = h_d$):

$$\mathbf{K}_{h,j}(x, v) = \mathbf{k}\left(\frac{x_1 - X_{1j}}{h}\right) \times \ldots \times \mathbf{k}\left(\frac{x_{d_x} - X_{d_x j}}{h}\right) \times \mathbf{k}\left(\frac{v - \widehat{V}_j}{h}\right)$$

and introduce the following shorthand notation for the first d_x elements:

$$\mathbf{K}_h(x - X_j) = \mathbf{k}\left(\frac{x_1 - X_{1j}}{h}\right) \times \ldots \times \mathbf{k}\left(\frac{x_{d_x} - X_{d_x j}}{h}\right)$$

To ensure uniform consistency and to bound the denominator away from zero, we introduce a nonrandom trimming function:

$$I_{xi} := I[x \in \mathcal{X}, V_i \in \mathcal{V}] \quad \text{and} \quad \widehat{I}_{xi} := I[x \in \mathcal{X}, \widehat{V}_i \in \mathcal{V}]$$

We refrain from using random trimming, but different trimming techniques might be used in practice. Finally, we also allow for random (right) censoring in the estimation of the conditional mean by using the so called "synthetic data" approach. As outlined in section 1, duration data is typcially subject to (random) right censoring. Instead of observing the mismeasured duration Y_j for each individual, we observe:

$$U_j = \min\{Y_j, C_j\} \quad \text{and} \quad \Delta_j = I\{Y_j \leq C_j\}$$

where C_j is the censoring time and Δ_j a censoring indicator. We assume $\{C_j, \Delta_j\}$ to be independent of the other model covariates. This assumption, albeit debatable in some settings, is standard in the literature and often justified in practice. In addition, define:

$$U_{jG} = \frac{U_j \Delta_j}{1 - G(U_j -)}$$

and

$$U_{j\widehat{G}} = \frac{U_j \Delta_j}{1 - \widehat{G}(U_j -)}$$

where $G(\cdot -)$ is the left-continuous distribution function of C_j and $\widehat{G}(\cdot -)$ the corresponding Kaplan-Meier estimator (Kaplan and Meier, 1958) with $\widehat{H}(\cdot -)$ the left-continuous empirical

distribution function of U_j:

$$\widehat{G}(c) = 1 - \prod_{i:C_i \leq c} \left(1 - \frac{\sum_{j=1}^{n} I[(1-\Delta_j) = 1, C_j \leq C_i]}{1 - \widehat{H}(U_i-)} \right)^{1-\Delta_i}$$

Replacing the partially unobserved Y_j by U_{jG}, Koul, Susarla, and van Ryzin (1981) showed that under condition B1 in the appendix:

$$\mathbb{E}[U_{jG}|X_j = x, V_j = v] = \mathbb{E}[Y_j|X_j = x, V_j = v] \tag{6}$$

Since U_{jG} is unobserved, we can replace it by $U_{j\widehat{G}}$ and estimate (6) as:

$$\widehat{\mu}(x, \widehat{V}_i) = \frac{\sum_{j=1}^{n} \widehat{I}_{xi} U_{j\widehat{G}} \mathbf{K}_{h,j}(x, \widehat{V}_i)}{\sum_{j=1}^{n} \widehat{I}_{xi} \mathbf{K}_{h,j}(x, \widehat{V}_i)} \tag{7}$$

while:

$$\widehat{\mu}(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mu}(x, \widehat{V}_i) \tag{8}$$

is the average of $\widehat{\mu}(x, \widehat{V}_i)$ over $\widehat{V}_i$. The last stage recovers the parameter vector β_0. As rank estimators only allow an identification of β_0 up to scale, we require a normalization of an arbitrary component of the parameter vector. Following standard procedures, we normalize the first component to one, i.e. $\beta(\theta) \equiv (1, \theta)$.[10] Thus, the third stage rank estimator is given by:

$$\beta(\widehat{\theta}) = \arg\max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{k \neq l} I[X_k \in \mathcal{X}] \times \widehat{\mu}(X_k) \times I[X_k' \beta(\theta) \geq X_l' \beta(\theta)] \tag{9}$$

where $\sum_{k \neq l}$ stands for the double sum $\sum_{k=1}^{n} \sum_{l>k}^{n}$ assuming that observations are in ascending order.[11] The form of (9) is almost identical to the two-stage rank estimator of Khan (2001) using a conditional mean instead of a conditional quantile function. We notice that for the above estimator to work we require that $\widehat{\mu}(X_k) > 0$ for every X_k in $\mathcal{X}$. Thus, if Y_j also takes on negative values, we require an upfront transformation of the data, e.g. $\overline{\overline{Y}}_j = Y_j - \min\{Y_1, \ldots, Y_n\}$, to ensure positivity.

[10] Accordingly, the true parameter vector is $\beta(\theta_0) \equiv (1, \theta_0)$.

[11] Summations appearing in the following that involve more than two indices will be defined according to the same logic.

2.3 Asymptotic Properties

This subsection considers the asymptotic properties of our estimation procedure. The probability limit of (9) evaluated at θ_0 is:

$$\beta(\theta_0) = \int I[X_k \in \mathcal{X}] \times \mu(X_k) \times I[X_k'\beta(\theta_0) \geq X_l'\beta(\theta_0)] dF_X(X_k, X_l) \tag{10}$$

where $F_X(\cdot,\cdot)$ in this case denotes the distribution function of X_k, X_l. Since the conditions for consistency, $\sqrt{n}$-consistency, and asymptotic normality are standard and rather lengthy (see Cavanagh and Sherman (1998) or Khan (2001) for details), we refer the reader to Appendix A.1, where we outline conditions B1 to B8 used in the theorems below together with a short discussion of non-standard assumptions. Notice that we employ a higher order kernel function in order to allow for a fairly large dimension of the covariate vector X_j. That is, with an increasing number of covariates used in the estimation of the conditional mean, we require a kernel function with an increasing number of moments equal to zero in order to control the bias.

Theorem 2. *Under conditions A1-A4, B1-B5, B7, and B8, we have:*

$$\widehat{\theta} \xrightarrow{p} \theta_0$$

The proof of Theorem 2 parallels the proof of Theorem 3.1 in Khan (2001). The main difference with respect to the latter, who uses a conditional quantile instead of a conditional mean estimator, is to show that replacing $\widehat{\mu}(X_k)$ by its probability limit $\mu(X_k)$ results in an error of smaller order for every $X_k \in \mathcal{X}$. Unlike Khan (2001), however, we also need to control for the estimated terms $\widehat{V}_j$, $U_{j\widehat{G}}$, and $\widehat{I}_j$. One difficulty arises as the $\widehat{V}_j$ also enter the indicator function $\widehat{I}_j$, which in turn prevents a Taylor expansion. We borrow an argument from Corradi, Distaso, and Swanson (2010) to show that this term can infact be bounded by an expression approaching zero at rate $\ln(n)^{\frac{1}{2}}/(nh^{d_z})^{\frac{1}{2}} \longrightarrow 0$. Together with the convergence rates of $U_{j\widehat{G}}$ and $\widehat{V}_j$, we obtain the overall rate:

$$\widehat{\mu}(x) - \mu(x) = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right) = o_p(1)$$

for every $x \in \mathcal{X}$.

Given consistency of $\widehat{\theta}$ for θ_0, we can replace the parameter space Θ by a shrinking set around θ_0 to establish $\sqrt{n}$-consistency and asymptotic normality using results of Sherman (1993). To simplify notation in the next theorem, we define the following expression (see Khan, 2001;

Sherman, 1993):

$$\psi_1(x,\theta) = \int \mu(x) \times I[x \in \mathcal{X}] I[x'\beta(\theta) > u'\beta(\theta)] - I[x'\beta_0 > u'\beta_0] dF_x(u) + \int \mu(u) \times I[u \in \mathcal{X}] I[u'\beta(\theta) > x'\beta(\theta)] - I[u'\beta_0 > x'\beta_0] dF_x(u) \tag{11}$$

Moreover, denote:

$$\psi_2(x,\theta) = \int I[x \in \mathcal{X}] I[x'\beta(\theta) > u'\beta(\theta)] dF_x(u) \tag{12}$$

Theorem 3. *Under conditions A1-A4 and B1-B8, we have:*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma = J^{-1}\Omega J^{-1}$ *with:*

$$J = \frac{1}{2}\mathbb{E}\big[\nabla_{\theta\theta'}\psi_1(X_k, \theta_0)\big]$$

The diagonal elements of the matrix Ω *are given by the sum of the following expressions:*

(i)

$$\Omega_0 = \int \Big(I_m(U_{mG} - \mu(X_m))\nabla_\theta \psi_2(X_m, \theta_0)\Big) \times \Big(I_m(U_{mG} - \mu(X_m))\nabla_\theta \psi_2(X_m, \theta_0)\Big)' dF_{U_G,X,V}(U_{mG}, X_m, V_m)$$

(ii) $\Omega_1 = E_1 \Phi_1 E_1'$ *with:*

$$\Phi_1 = \int V_i^2 dF_V(V_i)$$

and

$$E_1 = \Big(F_V^{(1)}\Big(a\Big) + F_V^{(1)}\Big(b\Big)\Big) \int U_{jG}\nabla_\theta \psi_2(X_k, \theta_0) dF_{U_G,X}(U_{jG}, X_k)$$

where a, b *are real numbers and* $F_V^{(1)}(\cdot)$ *denotes the first-order derivative of the distribution function* $F_V(\cdot)$ *of* V*.*

(iii) $\Omega_2 = E_2 \Phi_2 E_2'$ *with*

$$\Phi_2 = \Phi_1$$

and

$$E_2 = -\int I_i U_{jG} \nabla_\theta \psi_2(X_k, \theta_0) dF_{U_G,X,V}(U_{jG}, X_k, V_i)$$

(iv) $\Omega_3 = E_3 \Phi_3 E_3'$ *and*

$$\Phi_3 = \int_0^{\phi_Y} \mathbb{E}\Big[U_{1G} I[s < U_1]\Big] H_{t1}(s) \frac{dG(s)}{(1 - G(s-))}$$

and

$$E_3 = \int I_i \nabla_\theta \psi_2(X_k, \theta_0) dF_{X,V}(X_k, X_l, V_i)$$

where ϕ_Y *is defined in B1 and* $H_{t1}(s) = \mathbb{E}\Big[U_{1G} I[s < U_1]\Big] / \{(1 - F_Y(s-))(1 - G(s-))\}$.

The proof of Theorem (3) follows the proof of Theorem 3.2 in Khan (2001). We explicitly verify the conditions of Lemmata A.1 and A.2 therein, which establish $\sqrt{n}$-consistency and asymptotic normality, respectively. The main differences to Khan (2001) consist in the use of a conditional mean rather than a conditional quantile function and in the estimated first and second stage terms $\widehat{V}_j$, $\widehat{I}_j$, and $U_{j\widehat{G}}$, which complicate the asymptotic analysis in our case further. Both, the estimation of the conditional mean function as well as the estimated $\widehat{V}_j$, I_j, $U_{j\widehat{G}}$ yield the extra pieces Ω_0, Ω_1, Ω_2, and Ω_3 in the variance-covariance matrix Σ that differ from the expression derived by Khan (2001). The first step in the proof of the above theorem is to replace $\widehat{\mu}(X_k)$ in (9) by $\mu(X_k)$. The term involving $\mu(X_k)$ can be expanded to yield the gradient $J = \frac{1}{2}\mathbb{E}\big[\nabla_{\theta\theta'}\psi_1(X_k, \theta_0)\big]$ plus terms that are of order $o_p(n^{-1})$ once $\sqrt{n}$-consistency of $\|\widehat{\theta} - \theta_0\|$ has been established (notice that Lemmata B.1 and B.2 are verified concurrently and hence expressions shown to be of order $o_p(\|\widehat{\theta} - \theta_0\|/\sqrt{n})$ for instance automatically become $o_p(n^{-1})$ once $\|\widehat{\theta} - \theta_0\| = O_p(1/\sqrt{n})$ has been established via Lemma B.1). The second term containing the estimation error $(\widehat{\mu}(X_k) - \mu(X_k))$ on the other hand can be further expanded to give the different variance pieces plus terms that are again of order $o_p(n^{-1})$ on a set around θ_0 shrinking at rate $\sqrt{n}$.

2.4 Bootstrapping Confidence Intervals

The asymptotic variance depends on moments of the derivatives of the unknown functions $\psi_1(\cdot,\cdot)$ and $\psi_2(\cdot,\cdot)$, which can be estimated using either numerical derivatives (e.g. Sherman, 1993; Cavanagh and Sherman, 1998) or kernel-based methods (Abrevaya, 1999). However, since these moments may be difficult to estimate in practice, we propose to use the 'm out of n' bootstrapping procedure as an alternative to construct confidence intervals for our parameter estimates. The 'm out of n' bootstrapping procedure is a widely applicable methodology allowing to approximate the sampling distribution under fairly weak assumptions. Moreover, since other bootstrapping procedures fail to replicate U-statistic degeneracy (Arcones and Gine, 1992), they are not applicable in our setup.

The procedure works as follows: we sample $X_1^*, \ldots, X_m^*$ and $Z_1^*, \ldots, Z_m^*$ from our original sample of size n (with $m < n$) and obtain $\widehat{V}_1^*, \ldots, \widehat{V}_m^*$. We construct $1, \ldots, B$ of these bootstrap samples of size m. For each of these samples, we compute the bootstrap equivalent of our

estimator:

$$\beta(\theta^*) = \arg\max_{\theta\in\Theta} \frac{1}{m(m-1)} \sum_{k\neq l} I[X_k^* \in \mathcal{X}] \times \widehat{\mu}^*(X_k^*) \times I[X_k^{*\prime}\beta(\theta) \geq X_l^{*\prime}\beta(\theta)] \tag{13}$$

where

$$\widehat{\mu}^*(X_k^*) = \frac{1}{m} \sum_{i=1}^{m} \left\{ \frac{\sum_{j=1}^{m} \widehat{I}_{ki}^* U_{j\widehat{G}}^* \mathbf{K}_{h^*,j^*}(X_k^*, \widehat{V}_i^*)}{\sum_{j=1}^{m} \widehat{I}_{ki}^* \mathbf{K}_{h^*,j^*}(X_k^*, \widehat{V}_i^*)} \right\}$$

and the bandwidth sequence h^* is in lieu of h from Section 2.2 shrinking to zero at a rate depending on m (rather than n). Hence we obtain $\theta_1^*, \ldots, \theta_B^*$. Our aim is to construct a $1-\alpha$ confidence interval (CI) from the empirical bootstrap distribution. Thus, we need to recover standard errors from the bootstrap covariance matrix, which is given by:

$$\Sigma^* = \frac{m}{B} \sum_{i=1}^{B} \left(\theta_i^* - \frac{1}{B} \sum_{i=1}^{B} \theta_i^* \right) \left(\theta_i^* - \frac{1}{B} \sum_{i=1}^{B} \theta_i^* \right)'$$

The next theorem establishes that Σ^* is a consistent estimator for Σ:

Theorem 4. *Let $\mathbb{P}_*$ denote the probability distribution induced by the bootstrap sampling. Under assumptions A1-A4 and B1-B8 with h^* and m in place of h and n, respectively, and letting $m, n, \frac{n}{m} \longrightarrow \infty$, we have for all $\epsilon > 0$:*

$$\mathbb{P}\left(\omega : \mathbb{P}_*\left(\left| \Sigma^* - \Sigma \right| > \epsilon \right) \right) \longrightarrow 0$$

In order to prove the above theorem, we firstly verify that $\sqrt{m}(\theta^* - \widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta} - \theta_0)$ in a similar manner to before. However, since first order validity does not justify the use of the variance of the bootstrap distribution to consistently estimate the asymptotic variance (e.g. Goncalves and White, 2004), we need to show that uniform integrability holds as well. A sufficient condition for the latter is the existence of a slightly higher moment condition, which in turn ensures consistency of the bootstrap variance estimator.

3 Monte Carlo Simulations

To shed some light on the small sample properties of the estimator in 9, we conduct various Monte Carlo simulations in this section. The results are displayed in Table 1 and 2 of Appendix C. We consider four different measurement error designs, two without censoring (Design I and II) and two with censoring of different degree (Design III and IV). In all four cases, the data

generating process is linear and incorporates two independent variables X_{1j} and X_{2j}:

$$Y_j = X_{1j} + X_{2j}\theta_0 + \epsilon_j + \eta_j$$

with the coefficient of X_{1j} normalized to one and θ_0 set equal to .5. X_{2j} is chosen to be the endogenous variable, which is driven by the following first stage model:

$$X_{2j} = \alpha \cdot Z_j + V_j$$

with $\alpha = 1$. To ensure compactness of the covariate space, we follow Hahn, Hu, and Ridder (2008) and simulate Z_j and V_j from two uniform distributions $U[0,1]$ and $U[-1,1]$, repspectively. Notice that the chosen range of Z_j and V_j imply that V_j has full support for $0 \leq x_{2j} \leq 1$. Thus, all observations of X_{2j} are sampled from $[0,1]$ in the second stage. X_{1j} is drawn from a uniform distribution $U[1,2]$, while the idiosyncratic error and the measurement error differ according to the chosen design:

- Design I: $\epsilon_j \sim U[0,.5]$ and $\eta_j = \kappa \cdot V_j + \exp(\epsilon_j)$ with $\kappa = .6$.
- Design II: $\begin{pmatrix} \epsilon_j \\ \eta_j \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} .5 & .4 \\ .4 & .5 \end{pmatrix} \right)$.

For Design III and IV, we use the setup of Design I, but introduce an additional censoring variable C_j, which is sampled from a uniform distribution $U[2,10]$ (Design III) and $U[2,6]$ (Design IV), respectively. Notice that the support of C_j 'covers' the support of Y_j in both cases so that only the degree of (right) censoring varies.

We compare our estimation procedure (labelled RankCF) to its 'closest relative', the Monotone Rank Estimator (MRE) introduced by Cavanagh and Sherman (1998)[12], and the Maximum Rank Correlation Estimator (MRC) of Han (1987), which was the first estimator in the literature using a rank-type argument. A grid search method with 401 equispaced points on the interval $[-2,2]$ is employed to optimize the objective functions. The sample size varies from 100, 200, 400 to 600 observations. For every sample size, we conduct 401 replications. The displayed deviation measures are Mean Bias, Median Bias, Root Mean Square Error (RMSE), and Mean Absolute Deviation (MAD). They are constructed as averages over the number of replications. We employ a Gaussian kernel using Silverman's (1986) rule of thumb for the bandwidth selection.[13]

Starting with Table 1, one can observe that the multi-stage estimation procedure outperforms the MRE and the MRC for both designs at every sample size. In particular for smaller sample sizes, the RMSE and the MAD are fairly large, which is to be expected given the inconsistency

[12]We use the identity function as 'weighting' function of the dependent variable.

[13]Simulations with a second order Epanechnikov kernel provided similar results and are not reported here (available upon request).

of the MRE and the MRC and the additional estimation steps required for our methodology. However, while performance remains poor as sample size increases in the case of the former estimators, precision substantially improves for our estimation procedure.

Turning to the censoring setup in Table 2, we can again see that our procedure performs relatively well for both Designs III and IV, in particular excelling the MRE and the MRC in the case of 'light' censoring (Design III). As expected, all bias measures become fairly large once more pronounced censoring is introduced. Once again, however, we observe a substantial improvement for our method with the size of the sample growing, while the bias measures do not really change for the MRE and the MRC.

Overall, the results from this small simulation study indicate a good performance of our methodology in the linear setup under different forms of nonclassical measurement error and various degrees of censoring.

4 Empirical Illustration

In a recent study, Bricker and Engelhardt (2007) provided empirical evidence for nonclassical measurement error in annual earnings data from the Health and Retirement Study, which is a nationally representative longitudinal survey of the over 50 population in the US.[14] The researchers found a mean-reverting pattern in the data and a significant negative correlation between higher education and measurement error. Given the negative relationship between the amount of earnings and the sign of the measurement error, the latter finding is not too surprising as additional education is typically associated with higher earnings. Unlike in the paper of Bricker and Engelhardt (2007), we choose the 1998 wave, which also includes the 'War Babies' and the 'Children of the Great Depression' cohorts to broaden the age range in our data and to comply with the assumption of a continuous variable in the covariate vector. The sample is restricted to individuals with positive labour income during that year (i.e. no self-employed) and individuals that were the actual financial respondents of the household.[15] Moreover, to further ensure a certain degree of homogeneity, we only selected white individuals for our final dataset. The full support requirement in our assumption setup also meant that we had to exclude persons below the age of 50 and above 70, and those with less than 10 years of schooling. The final sample size comprised 2,753 observations.

For our earnings equation, we take the (natural) logarithm of annual labour income as dependent variable and consider gender, age (as a rough proxy for experience), and years of

[14] See the University of Michigan's webpage http://hrsonline.isr.umich.edu/index.php for a detailed description of the study and the data.

[15] Annual labour income comprises (i) regular wage or salary income, (ii) bonuses, tips, commissions, extra-pay from overtime, (iii) professional practice or trade earnings, and (iv) other income earned from a second job or while in the military reserves.

schooling as model covariates.[16] Using years of schooling as independent variable embeds the assumption of log earnings being a linear function of education, i.e. each additional year of education having the same proportional effect on annual earnings. Despite certain criticism in the literature about this hypothesis, Card (1999) points out that linearity often appears to fit the data surprisingly well and may thus not be easily rejected. We rule out the possibility of measurement error in the independent variables (which can certainly be put into question) and normalize the coefficient of gender to one. As instruments for the respondent's years of schooling we choose years of schooling of the mother and the father, respectively. These family background covariates are correlated with the schooling level of the individual, but unlikely to be related to the respondent's actual misreporting or his/ her ability.

The estimation results of our methodology are compared to the ones of the MRE and the MRC as well as a Least Squares (OLS), a Least Absolute Deviations (LAD), and a Two-Stage Least Squares (TSLS) estimator. The latter uses the mother's and the father's education as instrumental variables for the respondent's years of schooling and serves as an additional reference point for the education coefficient. Due to the discontinuous character of the objective function, we use a Nelder-Mead Simplex method to optimize the functions of the three rank estimators. As starting values for the initial simplex we choose an average of the OLS and LAD estimates.[17] To obtain a 95% confidence interval for the parameters, we conducted a 'm out of n' bootstrap with subsample size of 400 and 200 replications.

The education coefficient of our estimator in Table 3 of Appendix C differs from its competitors and falls substantially below their values hinting at an upwards bias in the education coefficient of the other estimators. This conjecture is backed by the TSLS estimator, which takes on the second-lowest value in the coefficient range (despite being potentially inconsistent due to its linearty restriction). The use of instrumental variables does naturally come at the price of larger confidence regions for the TSLS and our estimation procedure relative to the OLS and the other rank estimators (notice however that the point estimate of our procedure is still significantly different from zero at a 5% level). Moreover, we observe that the age coefficient (which was not found to be correlated with measurement error in the study of Bricker and Engelhardt (2007)) is substantially larger for our estimation procedure. Given inconsistency of the MRC and the MRE estimator and the linearity restrictions implicitly imposed by the OLS and the LAD estimator, this might once again be interpreted as a sign for measurement error bias in the coefficient estimate of age.

Summarizing this small illustrative example that looks at a log earnings equation with years of education, gender, and age as covariates, we find that point estimates provided by the estimation

[16]Notice that the use of three regressors plus an (estimated) control function requires the application of a third order kernel in theory. Since our simulation results from the previous section do however not display much sensitivity towards higher order kernel functions, we continue to employ the Gaussian kernel from Section 3 also in this empirical illustration.

[17]Noticce that the results were rather insensitive to small variations in the initial simplex.

procedure proposed in this paper differ quite substantially from those of its linear and non-linear competitors. This difference is particularly pronounced for the age coefficient. Together with evidence in the 1992 wave of the HRS for a mean-reverting non-classical measurement error in annual earnings that is correlated with education (Bricker and Engelhardt, 2007), this underlines the need to adjust for measurement error bias when examining the determinants of annual labour income of older workers in the HRS.

5 Conclusion

This paper proposes a multi-step procedure to identify and estimate the parameter vector of the monotone transformation model when the continuous dependent variable is subject to nonclassical measurement error. Empirical evidence examining duration and earnings data collected via survey questionnaires often suggests that such a measurement error represents the rule rather than the exception. Combining a modified control function approach with a rank-type argument, we show that it is possible to recover the aforementioned parameter vector consistently up to a location and size normalization. We derive the estimator's asymptotic properties and also demonstrate the methodology's good finite sample performance in a small Monte Carlo Study. Finally, we conclude with an empirical illustration investigating the effect of years of schooling on annual (log) earnings data from the Health and Retirement Study. We find substantially different point estimates using our estimation procedure (relative to other linear and nonlinear estimators) suggesting a potential measurement error problem when employing conventional estimators in this context.

Extensions of the present paper and topics for future research include the recovery of the unknown transformation function $m(\cdot)$ (the 'integrated baseline hazard' in survival analysis) and, in the duration context, the consideration of multiple spells. The latter in particular is non-trivial: despite suitable stationarity assumptions on the measurement error (similar to the ones used in Abrevaya (2000) for the idiosyncratic error terms), such an extension is more complex as 'fixed effects' estimators typically exploit 'intra-unit' variation rendering the integration over the support of the control function more difficult.

Appendix A

A.1 Assumptions

Let $\|\cdot\|$ denote the Euclidean norm and ∇_i the i-th order derivative of a function.

B1 C_j is i.i.d. and independent of Y_j. Moreover, C_j satisfies:

- **(i)** $\mathbb{P}[C_j \leq Y_j | Y_j = y, X_j = x, V_j = v] = \mathbb{P}[C_j \leq Y_j | Y_j = y]$.
- **(ii)** $G(\cdot)$ is continuous.
- **(iii)** $\phi_Y \leq \phi_C$
  with $\phi_Y = \inf\{t : F_Y(t) = 1\}$, $\phi_C = \inf\{t : G(t) = 1\}$, and $F_Y(t) = \mathbb{P}[Y_j \leq t]$, $G(t) = \mathbb{P}[C_j \leq t]$.
- **(iv)** When $\phi_Y < \phi_C$, $\limsup_{t \to \phi_Y} \left(\int_t^{\phi_Y} (1 - F_Y(s)) dG(s))^{1-\rho} / (1 - F_Y(t)) \right) < \infty$, for some $\frac{2}{5} < \rho < \frac{1}{2}$.
- **(v)** When $\phi_Y = \phi_C$, for some $0 \leq \varsigma < 1$, $(1 - G(t))^{\varsigma} = O((1 - F_Y(t-)))$ as $t \to \phi_Y$.
- **(vi)** Let $F_U(t) = \mathbb{P}[U_j \leq t]$ and $H(U_j) = \int_{-\infty}^{U_j} dG(s) / (\{1 - F_U(s)\}\{1 - G(s)\})$. Assume that:

$$\int U_j H^{\frac{1}{2}+\varepsilon}(U_j)[1 - G(U_j-)]^{-1} dF_{U,X,V}(U, X, V) < \infty$$

B2 The elements x in the support of X can be partitioned into subvectors of discrete $x^{(d)}$ and continuous $x^{(c)}$ components. Let $\mathcal{X}^{(d)}$ and $\mathcal{X}^{(c)}$ be the corresponding discrete and continuous parts of $\mathcal{X} \subset \mathcal{W}$. Assume that the conditional density (given $x^{(d)} \in \mathcal{X}^{(d)}$) on $\mathcal{W}$ is everywhere continuous and strictly bounded away from zero. Moreover, assume that $\mathcal{X}$ is not contained in any proper linear subspace of $\mathbb{R}^{d_x}$ and that the subset $\mathcal{X}_{(1)}$ of one component of the d_x-dimensional set $\mathcal{X} = \mathcal{X}^{(d)} \times \mathcal{X}^{(c)}$ contains the interval:

$$\left[\mu(x) - 3 \max_{x'_{(-1)}\theta} |x'_{(-1)}\theta| \quad ; \quad \mu(x) + 3 \max_{x'_{(-1)}\theta} |x'_{(-1)}\theta| \right]$$

for any $x \in \mathcal{X}$, where $x_{(-1)}$ denotes the remaining $(d_x - 1)$ dimensional component and the maximum is taken over $\mathcal{X}_{(-1)} \times \Theta$ with $\max_{x'_{(-1)}\theta} |x'_{(-1)}\theta| < \infty$.

B3 The multivariate kernel function $\mathbf{K}_h = k_h \times \ldots \times k_h$ with $\mathbf{K}_h : \mathbb{R}^d \longmapsto \mathbb{R}$ is symmetric, has compact support, and is differentiable (with bounded derivative). In addition, $\mathbf{K}_h(\cdot)$ satisfies (i) $\int \mathbf{K}_h(u) du = 1$, (ii) $\int \mathbf{K}_h(u) u^{\gamma} du = 0$ for $\gamma = 1, \ldots, r-1$, (iii) $\int \mathbf{K}_h(u) u^r du \neq 0$ and $\int \mathbf{K}_h(u) u^r du < \infty$, (iv) $\int |\mathbf{K}_h(u)| du < \infty$, and (v) $\int \mathbf{K}_h^2(u) du < \infty$.

B4 θ_0 lies in the interior of the parameter space Θ, a compact subset of $\mathbb{R}^{d-1}$.

B5 For any value $x^{(d)} \in \mathcal{X}^{(d)}$, assume that $\mu(\cdot)$ is twice differentiable in $x^{(c)}$. In addition, given $0 < \gamma \leq 1$ and $\delta_0 > 0$, for every $x_1^{(c)}, x_2^{(c)} \in \mathcal{X}^{(\rfloor)}$ and $i = 0, 1, 2$:

$$\left\| \nabla_i \mu(x_1^{(c)}, x^{(d)}) - \nabla_i \mu(x_2^{(c)}, x^{(d)}) \right\| \leq \delta_0 \left\| x_1^{(c)} - x_2^{(c)} \right\|^{\gamma}$$

where ∇_i denotes the order of derivative w.r.t $x^{(c)}$.

B6 Let $\psi_1(x, \theta)$ and $\psi_2(x, \theta)$ be defined as in (11) and (12):

- For each x in $\mathcal{X}$, $\psi_1(x, \cdot)$ is twice differentiable with second order Lipschitz derivative.
- $\mathbb{E}[\nabla_{\theta\theta'} \psi_1(\cdot, \theta_0)]$ is negative definite.
- For each $x \in \mathcal{X}$, $\psi_2(x, \cdot)$ is twice continuously differentiable in the second argument.
- $\mathbb{E}[\|U_G \nabla_{\theta} \psi_2(X, \theta) \nabla_i f_{X,V}(X, V)\|^2] < \infty$ and $\mathbb{E}[\|U_G \nabla_{\theta\theta'} \psi_2(X, \theta) \nabla_i f_{X,V}(X, V)\|^2] < \infty$ for all $\theta \in \Theta$ (where i denotes the order of derivative of $f_{X,V}(\cdot, \cdot)$ w.r.t. the first argument).

B7 Assume that $\mathbb{E}\left[V^2\right] < \infty$, $\mathbb{E}\left[\mu(x, V)^2\right] < \infty$, and $\mathbb{E}\left[\|U_G f_{X,V}(X, V)\|^2\right] < \infty$. Moreover, suppose that $F_V(\cdot)$ is continuously differentiable in its argument for every $V \in \mathcal{V}$.

B8 Let $d_z \leq d \leq r + \frac{1}{2}d_z$ (note that r is defined in B3). For $d < 3$, the bandwidth sequence h satisfies:

$$\left(\frac{1}{n}\right)^{\frac{1}{(3+d)}} < h < \left(\frac{\ln(n)}{n}\right)^{\frac{1}{2r+d_z}}$$

and for $d \geq 3$:

$$\left(\frac{1}{n}\right)^{\frac{1}{2d}} < h < \left(\frac{\ln(n)}{n}\right)^{\frac{1}{2r+d_z}}$$

.

Remark 1: B1 (i) together with the independence assumption of C_j are sufficient for the equality in (6). Condition (iii) ensures that we observe the entire distribution and, in combination with (iv) and (v), is relevant for the estimation of $G(U_j-))$ (see Lu and Cheng (2007) for details). The parameter ρ is determined by the "heaviness" of censoring, i.e. the smaller ρ, the fewer uncensored observations actually lie close to the "endpoint" ϕ_C. Finally, (vi) is a square integrability condition used in the proof of Theorem 3.

Remark 2: Assumption B2 extends condition A2 in the text, allowing also for discrete components in the parameter vector. The latter part of the condition ensures that identification is not lost by restricting ourselves to a compact subset of the support. That is, it is assumed that the set $\mathcal{X}_1$ of one regressor is sufficiently large (relative to the others), see Khan (2001) for details.

Remark 3: The requirement $d_z \leq d \leq r + \frac{1}{2}d_z$ of the bandwidth condition B8 allows to neglect the bias of the higher order kernel defined in B3. For a two dimensional instrument vector Z_j ($d_z = 2$) and a four dimensional covariate vector (three regressors plus the estimated control function $\widehat{V}_j$) as in the empirical illustration of section 4 for instance, we thus require the use of a third order kernel function to meet the above restriction and to be able to negelect the bias in the asymptotic distribution.

A.2 Proofs

Notice that it is implicitly understood that whenever $\mathbf{K}_{h,j}(\cdot,\cdot)$ is evaluated at $\widehat{V}_i$, we sum over $\widehat{V}_j$, while if the kernel function is evaluated at V_i, we sum over V_j. Moreover, we will suppress the dependency of $I[X_k \in \mathcal{X}, V_i \in \mathcal{V}]$ on X_k in the following and write the indicator function as I_i.

Proof of Lemma 1

By A1, we have that $Z_j \perp \eta_j, \epsilon_j, V_j$. Since Z_j is independent of both V_j and η_j, it follows by standard arguments:

$$Z_j \perp \eta_j | V_j$$

Moreover, since $X_j = g(Z_j) + v$ given $V_j = v$ is a function of Z_j only, this implies:

$$X_j \perp \eta_j | V_j$$

By identical arguments and using the fact that ϵ_j is independent of V_j, we can also establish that $X_j \perp \epsilon_j | V_j$. Hence, we obtain:

$$\begin{aligned}
\mathbb{E}[Y_j | X_j = x, V_j = v] &= \mathbb{E}[m(X_j'\beta_0 + \epsilon_j) + \eta_j | X_j = x, V_j = v] \\
&= \mathbb{E}[m(x'\beta_0 + \epsilon_j) | X_j = x, V_j = v] + \mathbb{E}[\eta_j | X_j = x, V_j = v] \\
&= \mathbb{E}[m(x'\beta_0 + \epsilon_j) | V_j = v] + \mathbb{E}[\eta_j | V_j = v]
\end{aligned}$$

where the last equality follows by conditional independence. Using the argument of Hahn, Hu, and Ridder (2008) or Imbens and Newey (2009), by condition A2 we can integrate over the marginal distribution of V and

apply iterated expectations to obtain:

$$\begin{aligned}\int \mathbb{E}[Y_j|X_j=x,V_j=v]f_V(v)dv &= \int \mathbb{E}[m(x'\beta_0+\epsilon_j)|V_j=v]f_V(v)dv + \int \mathbb{E}[\eta_j|V_j=v]f_V(v)dv \\ &= \mathbb{E}[m(x'\beta_0+\epsilon_j)] + \mathbb{E}[\eta_j]\end{aligned}$$

for each $x \in \mathcal{X}$. The result then follows by A3 and A4. That is, for two observations i, j $(i \neq j)$ with $x, \widetilde{x} \in \mathcal{X}$:

$$\mathbb{E}[m(x'\beta_0+\epsilon_i)] + \mathbb{E}[\eta_i] > \mathbb{E}[m(\widetilde{x}'\beta_0+\epsilon_j)] + \mathbb{E}[\eta_j] \quad \text{if} \quad x'\beta_0 > \widetilde{x}'\beta_0$$

Hence, the result follows. ■

Proof of Theorem 2

Using the same steps as in Theorem 3.1 of Khan (2001) and Lemma A1 below, the result follows instantly. ■

Lemma A1. Given B1 to B5, B7, and B8, we have that:

$$\widehat{\mu}(x) - \mu(x) = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right) = o_p(1)$$

for every $x \in \mathcal{X}$.

Proof of Lemma A1

Notice that:

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(x,\widehat{V}_i) - \mu(x)\right| \leq & \left|\frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(x,\widehat{V}_i) - \frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x,V_i)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mu}(x,V_i) - \mu(x)\right| \\ = & L_{1n} + L_{2n}\end{aligned} \tag{A-1}$$

where $\widetilde{\mu}(x,\cdot)$ is defined as (8) in the text with V_j, I_i, U_{jG} replacing $\widehat{V}_j, \widehat{I}_i, U_{j\widehat{G}}$. We start with L_{1n}, which can be decomposed as:

$$\begin{aligned}L_{1n} = & \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{s}_{x,V}(x,\widehat{V}_i) - \widetilde{s}_{x,V}(x,V_i)}{\widetilde{f}_{x,V}(x,V_i)} - \frac{\widetilde{f}_{x,V}(x,V_i) - \widehat{f}_{x,\widehat{V}}(x,V_i)}{\widetilde{f}_{x,V}(x,V_i)} \times \widehat{\mu}(x,\widehat{V}_i)\right\}\right| \\ = & L_{11n} + L_{12n}\end{aligned} \tag{A-2}$$

where

$$\widehat{s}_{x,\widehat{V}}(x,\widehat{V}_i) = \frac{1}{nh^d}\sum_{j=1}^{n}\widehat{I}_i U_{j\widehat{G}}\mathbf{K}_{h,j}(x,\widehat{V}_i) \tag{A-3}$$

and

$$\widehat{f}_{x,\widehat{V}}(x,\widehat{V}_i) = \frac{1}{nh^d}\sum_{j=1}^{n}\widehat{I}_i\mathbf{K}_{h,j}(x,\widehat{V}_i) \tag{A-4}$$

with $\widetilde{f}_{x,V}(x,V_i)$ and $\widetilde{s}_{x,V}(x,V_i)$ defined analoguously using U_{jG}, I_i, V_j, respectively (recall that $d = d_x + 1$). We examine L_{11n} first. This term can be further decomposed to tackle the random denominator:

$$L_{11n} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{s}_{x,V}(x,\widehat{V}_i) - \widetilde{s}_{x,V}(x,V_i)}{f_{x,V}(x,V_i)} + \left[\frac{1}{\widetilde{f}_{x,V}(x,V_i)} - \frac{1}{f_{x,V}(x,V_i)}\right](\widehat{s}_{x,V}(x,\widehat{V}_i) - \widetilde{s}_{x,V}(x,V_i))\right\}$$

By B3 and B8, the second term is of smaller order since $\sup_{x,V\in\mathcal{W}}\big|\widetilde{f}_{x,V}(x,V)-f_{x,V}(x,V)|=O_p((\ln(n)/nh^d)^{\frac{1}{2}})=o_p(1)$ with $f_{x,V}(x,V_i)$ denoting the true density evaluated at x,V_i. As for the first term, $f_{x,V}(x,V)$ is strictly bounded away from zero for all $x\in\mathcal{X}$ and $V\in\mathcal{V}$ by B2. A decomposition of the first term of L_{11n} yields:

$$\begin{aligned}&\Big|\frac{1}{n^2h^d}\sum_{i=1}^n\sum_{j=1}^n\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i-I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big|\\+&\Big|\frac{1}{n^2h^d}\sum_{i=1}^n\sum_{j=1}^n\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_h(x-X_j)\big\{k_h(\widehat{V}_i-\widehat{V}_j)-k_h(V_i-V_j)\big\}\Big|\\+&\Big|\frac{1}{n^2h^d}\sum_{i=1}^n\sum_{j=1}^n\frac{1}{f_{x,V}(x,V_i)}I_i\big(U_{j\widehat{G}}-U_{jG}\big)\mathbf{K}_{h,j}(x,V_i)\Big|\\+&o_p(1)\end{aligned}\tag{A-5}$$

where $o_p(1)$ captures terms of smaller order containing cross-products. Denote the first, second, and third term as L_{111n}, L_{112n}, and L_{113n}, respectively. We examine each of these terms separately, starting with L_{111n}. Notice that by A3, B2, B8, and standard arguments one can show that:

$$\max_{1\le j\le n}|\widehat{V}_j-V_j|=O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right)$$

Noting that $|\widehat{I}_i-I_i|=|I[a\le\widehat{V}_i\le b]-I[a\le V_i\le b]|$, we can use the same argument as in Lemma A3 of Newey, Powell, and Vella (1999) to show that for $\Delta_n=((\ln(n)/nh^{d_z})^{\frac{1}{2}})$ we have:

$$\begin{aligned}|\widehat{I}_i-I_i|=&|I[a\le V_i+(\widehat{V}_i-V_i)\le b]-I[a\le V_i\le b]|\\\le&\big(I[|V_i-a|\le\Delta_n]+I[|V_i-b|\le\Delta_n]\big)\end{aligned}$$

for $1\le i\le n$. Turning back to L_{111n}, this term can be expanded as:

$$\begin{aligned}L_{111n}\le&\Big|\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i-I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big|\\&+\Big|\frac{1}{n^2h^d}\sum_{i=1}^n\sum_{j=1}^n\Big\{\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i-I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\\&-\mathbb{E}\Big[f_{x,V}(x,V_i))\big(\widehat{I}_i-I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big\}\Big|\\=&M_1+M_2\end{aligned}$$

We consider M_1 first. Using the positivity of U_{jG}:

$$\begin{aligned}M_1\le&\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{x,V}(x,V_i)}|\widehat{I}_i-I_i|U_{jG}|\mathbf{K}_{h,j}(x,V_i)|\Big]\\\le&\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{x,V}(x,V_i)}\big(I[|V_i-a|\le\Delta_n]+I[|V_i-b|\le\Delta_n]\big)U_{jG}|\mathbf{K}_{h,j}(x,V_i)|\Big]\end{aligned}$$

We examine only the first term, the second one follows by identical arguments. Setting $a=0$ without loss of generality and letting $u_1=((x-X_j)/h)$, $u_2=((V_i-V_j)/h)$, and $f_V(\cdot)$ denote the density of V_i and V_j, after change of variables we obtain:

$$\begin{aligned}&\int\int\int\int_0^{\Delta_n}U_{jG}|\mathbf{K}_h(u_1)k_h(u_2)|\frac{f_{X,V}(x+u_1h,V_j+u_2h)}{f_{X,V}(x,V_j+u_2h)}f_{X,V}(x,V_j)dV_jdu_1du_2dF_{U_G}(U_G)\\=&\int\int_0^{\Delta_n}U_{jG}|\mathbf{K}_h(u_1)k_h(u_2)|f_{X,V}(x,V_j)dV_jdF_{U_G}(U_G)(1+O(h))\\=&O(\Delta_n)\end{aligned}$$

Next we consider M_2. The variance of this term is given by:

$$\mathbb{E}\Big[\Big(\frac{1}{n^2h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i) - \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big\}\Big)^2\Big] + O(\Delta_n^2)$$

The first expectation above can be dealt with in the same way as before. This yields:

$$\begin{aligned}&\mathbb{E}\Big[\Big(\frac{1}{n^2h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i) - \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)}\big(\widehat{I}_i - I_i\big)U_{jG}\mathbf{K}_{h,j}(x,V_i)\Big]\Big\}\Big)^2\Big]\\ =&O\Big(\frac{1}{n^2h^d}\Delta_n\Big) + O\Big(\Delta_n^2\Big)\end{aligned}$$

Using Chebychev's inequality and B8, $M_2 = o_p(\Delta_n)$, so the overall rate becomes:

$$L_{111n} = O_p\left(\left(\frac{\ln(n)}{nh^{d_z}}\right)^{\frac{1}{2}}\right)$$

Next we examine the second term of (A-5), L_{112n}. A mean value expansion around $(V_i - V_j)$ yields:

$$L_{112n} = \left|\frac{1}{n^2h^{d+1}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_h(x-X_j)k_h^{(1)}(\overline{V}_i - \overline{V}_j)\big((\widehat{V}_i - V_i) + (V_j - \widehat{V}_j)\big)\right|$$

where $\overline{V}_i, \overline{V}_j$ denote intermediate values and $k^{(1)}(\cdot)$ is the derivative of the kernel function w.r.t. its argument. We can rewrite the expression as:

$$\left|\frac{1}{n^2h^{d+1}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_h(x-X_j)k_h^{(1)}(\overline{V}_i - \overline{V}_j)\big(\big(\widehat{g}(Z_i) - g(Z_i)\big) + \big(g(Z_j) - \widehat{g}(Z_j)\big)\big)\right|$$

Since $(\widehat{g}(Z_i) - g(Z_i))$ and $(g(Z_j) - \widehat{g}(Z_j))$ are identical, we only examine the first term involving $(\widehat{g}(Z_i) - g(Z_i))$. Letting $\mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i) = \mathbf{K}_h(x-X_j) \times k_h^{(1)}(\overline{V}_i - \overline{V}_j)$, we can decompose the first term into:

$$\begin{aligned}L_{112n} \leq&\left|\frac{1}{nh^{d+1}}\sum_{i=1}^{n}\mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i)\Big] \times \Big(\widehat{g}(Z_i) - g(Z_i)\Big)\right| +\\ &\left|\frac{1}{n^2h^{d+1}}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i) - \mathbb{E}\Big[\frac{1}{f_{x,V}(x,V_i)}I_iU_{jG}\mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i)\Big]\Big)\Big(\widehat{g}(Z_i) - g(Z_i)\Big)\right|\\ =&N_{1n} + N_{2n}\end{aligned}$$

We start with N_{1n}. The expectation expression can be shown to be $O(1)$ using iterated expectations, change of variables, integration by parts, B1, B2, and B3. Moreover, since $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{g}(Z_i) - g(Z_i)$ converges in distribution (see Proof of Theorem 3), we have that $N_{1n} = O_p(n^{-\frac{1}{2}})$. The second term N_{2n} is of smaller order and can be shown to be $o_p(n^{-\frac{1}{2}})$ using similar arguments. Thus:

$$L_{112n} = O_p\left(\frac{1}{\sqrt{n}}\right)$$

It remains to consider L_{113n} of (A-5). Using the non-negativity the indicator function together with the

decomposition argument of Theorem 2 in Lu and Cheng (2007, p. 1915) for $\left|U_{j\widehat{G}} - U_{jG}\right|$ yields:

$$\begin{aligned} L_{113n} \leq & \frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{f_{x,V}(x, V_i)} I_i \left|U_{j\widehat{G}} - U_{jG}\right| \left|\mathbf{K}_{h,j}(x, V_i)\right| \\ & \leq \sup_{t \leq \phi_F} |\widehat{G}(t) - G(t)| \left[1 + \sup_{t \leq \max_j\{U_j\}} \frac{\left|\{\widehat{G}(t) - G(t)\}\right|}{\left|1 - \widehat{G}(t)\right|}\right] \times \\ & \frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \frac{U_j}{\{1 - G(U_j)\}^2} \right| \left| \frac{1}{f_{x,V}(x, V_i)} I_i \right| \left|\mathbf{K}_{h,j}(x, V_i)\right| \end{aligned}$$

By Srinivasan and Zhou (1994, p.199), we have that:

$$\sup_{t \leq \max_j\{U_j\}} \frac{\left|\{\widehat{G}(t) - G(t)\}\right|}{\left|1 - \widehat{G}(t)\right|} = O_p(1)$$

The term:

$$\frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \frac{U_j}{\{1 - G(U_j)\}^2} \right| \left| \frac{1}{f_{x,V}(x, V_i)} I_i \right| \left|\mathbf{K}_{h,j}(x, V_i)\right|$$

can again be dealt with in the same way as L_{112n} using B1 and B3 to show that it is $O_p(1)$. Finally, by conditions B1 and the result of Theorem 3.1 in Chen and Lo (1997):

$$\sup_{t \leq \phi_F} |\widehat{G}(t) - G(t)| = O_p(n^{-\rho})$$

for $\frac{2}{5} < \rho < \frac{1}{2}$ (where ρ in turn depends on the "heaviness" of censoring). Putting together these results, the rate of the piece is:

$$L_{113n} = O_p(n^{-\rho})$$

Hence, using B8, the convergence rate of L_{11n} becomes:

$$L_{11n} = O_p\left(\left(\frac{\ln(n)}{n h^{d_z}}\right)^{\frac{1}{2}}\right)$$

The same argument can be used to show that:

$$L_{12n} = O_p\left(\left(\frac{\ln(n)}{n h^{d_z}}\right)^{\frac{1}{2}}\right)$$

and hence the overall rate $O_p((\ln(n)/n h_{d_z})^{\frac{1}{2}})$ of L_{1n} follows.

Next we consider $L_{2n} = \left|\frac{1}{n}\sum_{i=1}^{n} \widetilde{\mu}(x, V_i) - \mu(x)\right|$. We examine the following decomposition:

$$L_{2n} \leq \left|\frac{1}{n}\sum_{i=1}^{n} \widetilde{\mu}(x, V_i) - \mu(x, V_i)\right| + \left|\frac{1}{n}\sum_{i=1}^{n} \mu(x, V_i) - \mu(x)\right|$$

where $\mu(x) = \mathbb{E}\Big[\mu(x, V_i)\Big]$. Since $\widetilde{\mu}(x, V_i)$ is a consistent estimator for $\mu(x, V_i)$ and $\mathbb{E}\Big[(\mu(x, V)^2\Big] < \infty$, we have:

$$\left|\frac{1}{n}\sum_{i=1}^{n} \widetilde{\mu}(x, V_i) - \mu(x, V_i)\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

Likewise, since $\mu(x, V_i)$ is continuous (and hence bounded) on $\mathcal{W}$ and B7, we have that:

$$\left|\frac{1}{n}\sum_{i=1}^{n} \mu(x, V_i) - \mu(x)\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

■

Proof of Theorem 3

Let $A_{lk}(\theta) = I[X_k \in \mathcal{X}]\{I[X_k'\beta(\theta) > X_l'\beta(\theta)] - I[X_k'\beta(\theta_0) > X_l'\beta(\theta_0)]\}$. Since the second term involving $\beta(\theta_0)$ does not affect maximization, we note that $\widehat{\theta}$ still maximizes:

$$Q_n(\theta) = \frac{1}{n(n-1)} \sum_{k \neq l} \widehat{\mu}(X_k) A_{lk}(\theta) \tag{A-6}$$

and θ_0 its corresponding probability limit:

$$Q(\theta) = \mathbb{E}\Big[\mu(X_k) A_{lk}(\theta)\Big] \tag{A-7}$$

Notice in addition that by the normalizations in (A-6) and (A-7), we have that $Q_n(\theta_0) = Q(\theta_0) = 0$. We expand $Q_n(\theta)$ around the true $\mu(X_k)$ yielding:

$$\begin{aligned} &\frac{1}{n(n-1)} \sum_{k \neq l} \mu(X_k) A_{lk}(\theta) + \frac{1}{n(n-1)} \sum_{k \neq l} \Big(\widehat{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta) \\ =& S_{1n} + S_{2n} \end{aligned}$$

In the following, we proceed by examining S_{1n} and S_{2n} in turn, starting with S_{1n}. Identical arguments to the proof of Lemma A.3 in Khan (2001) can be used to show that S_{1n} yields the gradient term plus terms that are of order $o_p(n^{-1})$ once $\sqrt{n}$-consistency of $\|\theta - \theta_0\|$ has been established. That is:

$$S_{1n} = (\theta - \theta_0)' J (\theta - \theta_0) + o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big) + o_p\Big(\|\theta - \theta_0\|^2\Big) + o_p\Big(\frac{1}{n}\Big)$$

with

$$J = \frac{1}{2}\, \mathbb{E}\Big[\nabla_{\theta\theta'} \psi_1(X_k, \theta_0)\Big]$$

and $\psi_1(x, \theta)$ defined in (11) of section 2.3.

S_{2n} on the other hand can be further expanded to give:

$$\begin{aligned} &\frac{1}{n(n-1)} \sum_{k \neq l} \Big(\widetilde{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta) + \frac{1}{n(n-1)} \sum_{k \neq l} \Big(\widehat{\mu}(X_k) - \widetilde{\mu}(X_k)\Big) A_{lk}(\theta) \\ =& S_{21n} + S_{22n} \end{aligned}$$

where $\widetilde{\mu}(x)$ is defined analoguously to $\widehat{\mu}(x)$ using the true U_{jG}, I_i, V_j. S_{21n} and S_{22n} determine the components of the variance. They can be tackled through Lemma B3 and Lemmata B4 to B6, respectively: using the result of Lemma B3 below, $S_{21n} = (\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{0n} + o_p\big(\|\theta - \theta_0\|/\sqrt{n}\big)$, where W_{0n} is a sum of zero mean vector random variables that converges in distribution to a random vector defined in Lemma B3. It remains to examine S_{22n}, which can be expanded as in the proof of Theorem 2:

$$\begin{aligned} S_{22n} =& \frac{1}{n(n-1)} \sum_{k \neq l} \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\widehat{s}_{X,V}(X_k, \widehat{V}_i) - \widetilde{s}_{X,V}(X_k, V_i)}{\widetilde{f}_{X,V}(X_k, V_i)} + \frac{\widetilde{f}_{X,V}(X_k, V_i) - \widehat{f}_{X,V}(X_k, \widehat{V}_i)}{\widetilde{f}_{X,V}(X_k, V_i)} \widehat{\mu}(X_k, \widehat{V}_i) \right\} A_{lk}(\theta) \\ =& S_{22n}^{(1)} + S_{22n}^{(2)} \end{aligned}$$

where $\widehat{s}_{X,V}(\cdot,\cdot)$ and $\widehat{f}_{X,V}(\cdot,\cdot)$ are defined in (A-3) and (A-4), respectively, and $\widetilde{s}_{X,V}(\cdot,\cdot)$ and $\widetilde{f}_{X,V}(\cdot,\cdot)$ follow

accordingly. We start with $S_{22n}^{(1)}$, which can be further decomposed into:

$$\begin{aligned}
S_{22n}^{(1)} =& \frac{1}{n(n-1)} \sum_{k \neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum_{j=1}^{n} \big(\widehat{I}_i - I_i\big) U_{jG} \mathbf{K}_{h,j}(X_k, V_i)}{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)} \right\} \times A_{lk}(\theta) \\
&+ \frac{1}{n(n-1)} \sum_{k \neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum_{j=1}^{n} I_i U_{jG} \big(\mathbf{K}_{h,j}(X_k, \widehat{V}_i) - \mathbf{K}_{h,j}(X_k, V_i)\big)}{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)} \right\} \times A_{lk}(\theta) \\
&+ \frac{1}{n(n-1)} \sum_{k \neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \big(U_{j\widehat{G}} - U_{jG}\big) \mathbf{K}_{h,j}(X_k, V_i)}{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)} \right\} \times A_{lk}(\theta) \\
&+ o_p(1) \\
=& S_{221n} + S_{222n} + S_{223n} + o_p(1)
\end{aligned} \tag{A-8}$$

where the $o_p(1)$ term contains cross-products of smaller order. We examine each of the three terms seperately starting with S_{221n}, which by Lemma B4 is equal to:

$$(\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{1n} + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \right) + o_p\left(\frac{1}{n} \right)$$

where W_{1n} is again defined in Lemma B4 below. Likewise, for S_{222n} and S_{223n}, we can apply Lemma B5 and B6 to obtain:

$$S_{222n} = (\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{2n} + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \right) + o_p\left(\frac{1}{n} \right)$$

and

$$S_{223n} = (\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{3n} + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \right) + o_p\left(\frac{1}{n} \right)$$

with W_{2n} and W_{3n} being again sums of zero mean vector random variables that converge to a normal distribution defined in Lemma B5 and B6, respectively. Next we consider $S_{22n}^{(2)}$. A similar decomposition as for $S_{22n}^{(1)}$ and arguments as in Lemmata B4 and B5 can be used to show that the limiting distribution of this term is the same as that of S_{221n} and S_{222n}.

Taking these decompositions of S_{1n} and S_{2n} together and using B5, Lemma B1 and B2 below become directly applicable establishing $\sqrt{n}$-consistency and asymptotic normality. Notice that for (i) of Lemma B1, b_n can be set to be $o(1)$ by the consistency result of Theorem 2. (ii) of the same lemma is satisfied by B2 and B5 in combination with a second order Taylor expansion of $Q(\theta)$ in (A-7) around θ_0: $Q(\theta) = \frac{1}{2}(\theta - \theta_0)' \nabla_{\theta\theta'} Q(\overline{\theta})(\theta - \theta_0) \leq -\kappa \|\theta - \theta_0\|^2$ for some constant κ and $\overline{\theta} \in \Theta$. ■

The following two lemmata are from Theorem 3.2 of Khan (2001) (we adapt notation of the original paper to our setup).

Lemma B1. (Lemma A.1 of Khan (2001)) Let $\widehat{\theta}$ maximize $Q_n(\theta)$ in (A-6) and θ_0 maximizes $Q(\theta)$ in (A-7). Let $b_n, l_n \longrightarrow 0$ as $n \longrightarrow \infty$. If:

- **(i)** $\widehat{\theta} - \theta_0 = O_p(b_n)$
- **(ii)** there exists a neighbourhood $\mathcal{N}$ of 0 and a positive constant κ for which:

$$Q(\theta) \leq -\kappa \|\theta - \theta_0\|^2$$

for all θ in $\mathcal{N}$,

(iii) uniformly over $O_p(b_n)$ neighbourhoods of 0,

$$Q_n(\theta) = Q(\theta) + O_p(\|\theta - \theta_0\|/\sqrt{n}) + o_p(\|\theta - \theta_0\|^2) + O_p(l_n) \tag{A-9}$$

then:

$$\|\widehat{\theta} - \theta_0\| = O_p\Big(\max\{l_n^{\frac{1}{2}}, 1/\sqrt{n}\}\Big)$$

Lemma B2. (Lemma A.2 of Khan (2001)) Suppose $\widehat{\theta}$ is $\sqrt{n}$-consistent for θ_0, an interior point of Θ. Suppose also that uniformly over $O_p(1/\sqrt{n})$ neighbourhoods of 0:

$$Q_n(\theta) = (\theta - \theta_0)' J (\theta - \theta_0) + \frac{1}{\sqrt{n}}(\theta - \theta_0)' W_n + o_p(1/n) \tag{A-10}$$

where J is a negative definite matrix, and W_n converges in distribution to a $N(0, \Sigma)$ random vector. Then

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, J^{-1}\Sigma J^{-1})$$

.

Lemma B3. Under assumptions A1-A4, B1-B6, and B8, the term $\frac{1}{n(n-1)} \sum_{k \neq l} \Big(\widetilde{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta)$ is equal to:

$$(\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{0n} + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right) = (\theta - \theta_0)' \frac{1}{n} \sum_{m=1}^{n} I_m (U_{mG} - \mu(X_m)) \nabla_\theta \psi_2(X_k = X_m, \theta_0) + o_p\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right)$$

where

$$\frac{1}{\sqrt{n}} \sum_{m=1}^{n} I_m (U_{mG} - \mu(X_m)) \nabla_\theta \psi_2(X_k = X_m, \theta_0) \xrightarrow{d} N(0, \Omega_0)$$

with

$$\Omega_0 = \int \Big(I_m (U_{mG} - \mu(X_m)) \nabla_\theta \psi_2(X_m, \theta_0)\Big) \times \Big(I_m (U_{mG} - \mu(X_m)) \nabla_\theta \psi_2(X_m, \theta_0)\Big)' dF_{U_G, X, V}(U_{mG}, X_m, V_m)$$

where $\psi_2(\cdot, \cdot)$ is defined in (12) of section 2.3.

Proof of Lemma B3

Notice that $\frac{1}{n(n-1)} \sum_{k \neq l} \Big(\widetilde{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta)$ can be rewritten as:

$$\frac{1}{n(n-1)} \sum_{k \neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \big(U_{jG} - \mu(X_k)\big) \mathbf{K}_{h,j}(X_k, V_i)}{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)} \right\} A_{lk}(\theta)$$

$$= \frac{1}{n(n-1)} \sum_{k \neq l} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{1}{nh^d} \sum_{j=1}^{n} I_i \big(U_{jG} - \mu(X_k)\big) \mathbf{K}_{h,j}(X_k, V_i)}{f_{X,V}(X_k, V_i)} \right\} A_{lk}(\theta) + o_p(1)$$

where $\widehat{f}_{X,V}(X_k, V_i) = \frac{1}{nh^d} \sum_{j=1}^{n} I_i \mathbf{K}_{h,j}(X_k, V_i)$ and the $o_p(1)$ term follows as in the proof of Theorem 2 by B3 and the bandwidth condition B8. As for the first term, $f_{X,V}(X, V)$ is strictly bounded away from zero for every

$X, V \in \mathcal{W}$ by B2 and can be restated as:

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} \frac{1}{h^d} \frac{1}{f_{X,V}(X_k, V_i)} I_i\big(U_{jG} - \mu(X_k)\big)\mathbf{K}_{h,j}(X_k, V_i) A_{lk}(\theta) \tag{A-11}$$

where omitting terms with $k = l = i = j$ results in an error of magnitude $o_p(\|\theta - \theta_0\|/nh^d)$. The expression in (A-11) is a fourth order U-statistic for each $\theta \in \Theta$. Letting $\xi_k = \{I_k, U_{kG}, X_k, V_k\}$ (ξ_l, ξ_i, ξ_j are defined accordingly), (A-11) is:

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} q_n(\xi_i, \xi_j, \xi_k, \xi_l; \theta) \tag{A-12}$$

where $q_n(\cdot,\cdot,\cdot,\cdot;\theta) = \frac{1}{h^d} I_i\big(U_{jG} - \mu(X_k)\big)\mathbf{K}_{h,j}(X_k, V_i) A_{lk}(\theta)$ is the 'kernel' function of the U-statistic. Using iterated expectations repeatedly, change of variables, together with B1, B3, B5, and B6 one can show that $q_n(\cdot,\cdot,\cdot,\cdot)$ is degenerate in ξ_k, ξ_l, and ξ_i for each $\theta \in \Theta$ since the expectation of $(U_{jG} - \mu(X_k))$ conditional on X_k is zero. By iterated expectations, this in turn implies that $\mathbb{E}\Big[q_n(\xi_k, \xi_l, \xi_i, \xi_j, \theta)\Big] = 0$. By contrast, after change of variables, iterated expectations, and dominated convergence, the term $\mathbb{E}\Big[q_n(\xi_k, \xi_l, \xi_i, \xi_j, \theta)\Big|\xi_j\Big]$ yields:

$$\begin{aligned}
\lim_{n \longrightarrow \infty} \mathbb{E}\Big[q_n(\xi_k, \xi_l, \xi_i, \xi_j, \theta)\Big|\xi_j\Big] =& \mathbb{E}\Big[q(\xi_k, \xi_l, \xi_i, \xi_j, \theta)\Big|\xi_j\Big] \\
=& \Big(U_{jG} - \mu(X_j)\Big) I_j \, \mathbb{E}\Big[A_{lk}(\theta)\Big|X_k = X_j\Big] \\
=& (\theta - \theta_0)' \Big(U_{jG} - \mu(X_j)\Big) I_j \nabla_\theta \psi_2(X_j, \theta_0) + O(\|\theta - \theta_0\|^2)
\end{aligned}$$

where the last equality follows by a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k = X_j\Big]$ around θ_0. Applying the Hoeffding decomposition to the degenerate fourth order U-process in (A-12) (Serfling, 1980) and noting that, by Lemma A.6 in Khan (2001) and the arguments used therein, all terms except the leading term are of order $o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big)$, yields:

$$\frac{1}{n(n-1)(n-2)(n-3)} \sum_{k \neq l \neq i \neq j} q_n(\xi_i, \xi_j, \xi_k, \xi_l; \theta) = \frac{1}{n} \sum_{m=1}^{n} \mathbb{E}\Big[q(\xi_k, \xi_l, \xi_i, \xi_j, \theta)\Big|\xi_j = \xi_m\Big] + o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big)$$

and hence:

$$(\theta - \theta_0)' \frac{1}{n} \sum_{m=1}^{n} \big(U_{mG} - \mu(X_m)\big) I_m \nabla_\theta \psi_2(X_k = X_m, \theta_0) + o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big)$$

The expression:

$$\frac{1}{n} \sum_{m=1}^{n} \big(U_{mG} - \mu(X_m)\big) I_m \nabla_\theta \psi_2(X_k = X_m, \theta_0) \tag{A-13}$$

is a sum of zero mean random variables. Applying Lindberg Levy's Central Limit Theorem (CLT) yields the result of the lemma. ■

Lemma B4. Under assumptions A1-A4, B1-B8, the term S_{221n} defined in the proof of Theorem 3 is equal to:

$$\begin{aligned}
&(\theta - \theta_0)' \frac{1}{\sqrt{n}} W_{1n} + o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big) + o_p\Big(\frac{1}{n}\Big) \\
=&(\theta - \theta_0)' \frac{1}{n} \sum_{m=1}^{n} \Big(\widehat{g}(Z_m) - g(Z_m)\Big)\Big(F_V^{(1)}\Big(a\Big) + F_V^{(1)}\Big(b\Big)\Big) \int U_{jG} \nabla_\theta \psi_2(X_k, \theta_0) dF_{U_G, X, V}(U_G, X_k, V_i)\Big) \\
&+ o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big) + o_p\Big(\frac{1}{n}\Big)
\end{aligned}$$

where

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(F_V^{(1)}\Big(\overline{V}_a\Big)+F_V^{(1)}\Big(\overline{V}_b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X}(U_G,X_k)\Big)\xrightarrow{d}N\Big(0,\Omega_1\Big)$$

with $\Omega_1 = E_1\Phi_1 E_1^{'}$:

$$\Phi_1 = \int V_i^2 dF_V(V_i)$$

and

$$E_1 = \Big(F_V^{(1)}\Big(a\Big)+F_V^{(1)}\Big(b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_{jG},X_k,V_i)$$

where $F_{U_G,X,V}(\cdot,\cdot)$ denotes the joint distribution function of U_{jG} and X_k.

Proof of Lemma B4

As before, we start by replacing $\widehat{f}_{X,V}(X_k,V_i)$ with the true density $f_{X,V}(X_k,V_i)$ using B2, B3, and B8. Moreover, notice that $I\{a\le\widehat{V}_j\le b\}-I\{a\le V_i\le b\} = I\{\widehat{V}_i\le b\}+I\{\widehat{V}_i\ge a\}-I\{V_i\le b\}-I\{V_i\ge a\} = \big(I\{\widehat{V}_i\le b\}-I\{V_i\le b\}\big)+\big(I\{\widehat{V}_i\ge a\}-I\{V_i\ge a\}\big)$. We focus on $\big(I\{\widehat{V}_i\le b\}-I\{V_i\le b\}\big)$, the other term will follow by an identical argument. Let $F_V(b)$ denote the distribution function of V_i evaluated at b and $\mathbf{B}_{ijkl}(\theta) = f_{X,V}^{-1}(X_k,V_i)U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)$. Then we can decompose S_{221n} as follows:

$$\begin{aligned}
&\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\big(I\{V_i\le b+(\widehat{V}_i-V_i)\}-I\{V_i\le b\}\big)\frac{1}{h^d}\mathbf{B}_{ijkl}(\theta)\\
=&\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big\{I\{V_i\le b+(\widehat{V}_i-V_i)\}-F_V\Big(b+(\widehat{V}_i-V_i)\Big)\\
&-I\{V_i\le b\}+F_V\Big(b\Big)\Big\}\times\frac{1}{h^d}\mathbf{B}_{ijkl}(\theta)\\
&+\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(F_V\Big(b+(\widehat{V}_i-V_i)\Big)+F_V\Big(b\Big)\Big)\frac{1}{h^d}\mathbf{B}_{ijkl}(\theta)\\
=&T_{1n}(\theta)+T_{2n}(\theta)
\end{aligned}$$

We start with $T_{1n}(\theta)$. We examine the term involving $F_V\Big(b\Big)-I\{V_i\le b\}$, the term with $I\{V_i\le b+(\widehat{V}_i-V_i)\}-F_V\Big(b+(\widehat{V}_i-V_i)\Big)$ follows by the same argument. Adding and subtracting $\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijkl}(\theta)\Big]$ yields:

$$\begin{aligned}
T_{1n}(\theta) =&\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(F_V\Big(b\Big)-I\{V_i\le b\}\Big)\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]\\
&+\frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(F_V\Big(b\Big)-I\{V_i\le b\}\Big)\times\\
&\Big(\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)-\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]\Big)\\
=&T_{11n}(\theta)+T_{12n}(\theta)
\end{aligned}$$

We start with the first piece, which can be simplified since no term depends on k,l, or j:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(F_V\Big(b\Big)-I\{V_i\le b\}\Big)\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]$$

Since $\mathbb{E}\Big[F_V\Big(b\Big)-I\{V_i\le b\}\Big]=0$, notice that by change of variables, iterated expectations, and a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k=X_j\Big]$ around θ_0, the variance of T_{11n} is $\mathbb{E}\Big[\Big(T_{11n}(\theta)\Big)^2\Big]=O\Big(\frac{\|\theta-\theta_0\|^2}{n}\Big)$. Thus,

using Chebychev's inequality, we have that $T_{11n}(\theta) = o_p(n^{-1})$.

Next, we consider $T_{12n}(\theta)$. To derive an upper bound for the convergence rate of $T_{12n}(\theta)$ via Rosenthal's inequality, we first examine:

$$\begin{aligned}
&\mathbb{E}\Bigg[\Big(I\{V_i \leq b\} - F_V\Big(b\Big)\Big)^2 \Bigg\{\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]^2 \\
&+ 2\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta) \times \mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}U_{jG}\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big] \\
&+ \frac{1}{h^{2d}}\frac{1}{f_{X,V}(X_k,V_i)^2}U_{jG}^2\mathbf{K}_{h,j}^2(X_k,V_i)A_{lk}^2(\theta)\Bigg\}\Bigg] \\
=&T_{121}(\theta) + T_{122}(\theta) + T_{123}(\theta)
\end{aligned}$$

We start with $T_{121}(\theta)$. Using change of variables, iterated expectations, a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)\Big|X_k = X_j\Big]$ around θ_0, B3, B7, and the boundedness of the indicator function, we have that $T_{121}(\theta) = O(\|\theta-\theta_0\|^2) + o(\|\theta-\theta_0\|^2)$. By the same line of argument, the same rates can be obtained for $T_{122}(\theta)$. Using again $u_1 = (X_k - X_j)/h$, $u_2 = (V_i - V_j)/h$, boundedness of the indicator function, B2, B3, B6, B7, and the equality $A_{lk}^2(\theta) = |A_{lk}(\theta)|$, $T_{123}(\theta)$ on the other hand is given by:

$$\begin{aligned}
T_{123}(\theta) =& \int \Big(I\{V_i \leq b\} - F_V\Big(b\Big)\Big)^2 \frac{1}{h^d}\frac{1}{f_{X,V}(X_j+hu,V_j+hu_2)^2}U_{jG}^2\mathbf{K}_{h,j}^2(X_j+hu_1,V_j+hu_2) \\
&\times \Big|(\theta-\theta_0)'\nabla_\theta\psi_2(X_j+hu_1,\theta_0)\Big| f_{X,V}(X_j,V_j)f_{X,V}(X_j+hu_1,V_j+hu_2) \\
&\times dx_j du_1 dv_j du_2 dU_G + o\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big) \\
=&O\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big)(1+h) + o\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big)
\end{aligned}$$

Moreover, using identical arguments:

$$\mathbb{E}\Bigg[\Bigg|\Big(F_V\Big(b\Big) - I\{V_i \leq b\}\Big)\Big(\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta) - \mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]\Big)\Bigg|^\kappa\Bigg] = O(\|\theta-\theta_0\|h^{-d(\kappa-1)}) + o(\|\theta-\theta_0\|h^{-d(\kappa-1)})$$

for $\kappa \geq 1$. Applying Rosenthal's inequality yields:

$$\begin{aligned}
\mathbb{E}\Big[\Big(T_{12n}(\theta)\Big)^{2\kappa}\Big] \leq& n^{-8\kappa}\Xi_\kappa\Big(\|\theta-\theta_0\|n^{4\kappa}h^{-d\kappa} + \|\theta-\theta_0\|n^4h^{-2\kappa d+d}\Big) \\
=&O\Big(\|\theta-\theta_0\|n^{-4\kappa}h^{-d\kappa}\Big) + O\Big(\|\theta-\theta_0\|n^{-8\kappa+4}h^{-2\kappa d+d)}\Big)
\end{aligned}$$

with Ξ_κ some positive constant. Using B6, we obtain the following rates for $\kappa = 1$: $O(\|\theta-\theta_0\|n^{-4}h^{-d}) + O(\|\theta-\theta_0\|n^{-4}h^{-2d+d}) = O(\|\theta-\theta_0\|n^{-4}h^{-d})$. By the bandwidth conditions in B8, Markov's inequality thus implies $T_{12n}(\theta) = o_p(n^{-1})$.

Next, we consider $T_{2n}(\theta)$, which can again be decomposed as:

$$\begin{aligned}
T_{2n}(\theta) =& \frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(F_V\Big(b+(\widehat{V}_i - V_i)\Big) + F_V\Big(b\Big)\Big)\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] \\
&+ \frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(F_V\Big(b+(\widehat{V}_i - V_i)\Big) + F_V\Big(b\Big)\Big) \\
&\times \Big(\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta) - \mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big]\Big) \\
=&T_{21n}(\theta) + T_{22n}(\theta)
\end{aligned}$$

We start with T_{21n}. Using B7, a mean value expansion around $(\widehat{V}_i - V_i) = 0$, and a simplification (since T_{21n}

only depends on j) yield:

$$\frac{1}{n}\sum_{i=1}^{n} F_V^{(1)}\Big(\overline{V}_b\Big)\Big(\widehat{V}_i - V_i\Big)\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] = \frac{1}{n}\sum_{i=1}^{n} F_V^{(1)}\Big(\overline{V}_b\Big)\Big(\widehat{g}(Z_i) - g(Z_i)\Big)\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] \tag{A-14}$$

where $\overline{V}_b \in [b, b+(\widehat{V}_i - V_i)]$ and $F_V^{(1)}$ denotes the first derivative w.r.t. its argument. Using iterated expectations, change of variables, and a second order Taylor expansion of $\mathbb{E}\Big[A_{lk}(\theta)|X_k\Big]$ around θ_0, the expectation expression in (A-14) yields:

$$\mathbb{E}\Big[\frac{1}{h^d}\mathbf{B}_{ijlk}(\theta)\Big] = (\theta-\theta_0)^{'}\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X}(U_G,X_k)(1+O(h)) + o(\|\theta-\theta_0\|^2)$$

For the random component in (A-14), recall that $\widehat{g}(Z_i) = (\sum_{j=1}^{n} X_{1j}\mathbf{k}_h(Z_i - Z_j))/\sum_{j=1}^{n}\mathbf{k}_h(Z_i - Z_j))$. We examine the following standard decomposition:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{g}(Z_i) - g(Z_i)\widehat{f}_Z(Z_i)}{\widehat{f}_Z(Z_i)} \\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\Big(\widehat{g}(Z_i) - g(Z_i)\widehat{f}_Z(Z_i)\Big)}{f_Z(Z_i)} + \left(\frac{f_Z(Z_i) - \widehat{f}_Z(Z_i)}{f_Z(Z_i)\widehat{f}_Z(Z_i)}\right)\times\Big(\widehat{g}(Z_i) - g(Z_i)\widehat{f}_Z(Z_i)\Big)\right\}
\end{aligned}$$

Since $\sup_{Z\in\mathcal{W}}|f_Z(Z_i) - \widehat{f}_Z(Z_i)| = o_p((\ln(n)/nh^d)^{\frac{1}{2}}) = o_p(1)$ by B3 and B8, the second term is of smaller order than the first one and will hence be neglected. Moreover, since $X_{1i} = g(Z_i) + V_i$, observe that the first term is can be restated as:

$$\frac{1}{n^2h}\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\frac{g(Z_j) - g(Z_i)}{f_Z(Z_i)}\right)\mathbf{k}_h(Z_i - Z_j) + \frac{1}{n^2h}\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{V_j}{f_Z(Z_i)}\mathbf{k}_h(Z_i - Z_j)$$

Now, notice that omitting observations with $i = j$ results in a negligible error of order $o_p((nh)^{-1})$, while $\frac{1}{2}(V_j f_Z^{-1}(Z_i) - V_i f_Z^{-1}(Z_j))\mathbf{k}_h(Z_i - Z_j)$ is the 'symmetrized' version of the second term and:

$$\frac{1}{2h}\left(\frac{(g(Z_i) - g(Z_j))}{f_Z(Z_j)} - \frac{(g(Z_j) - g(Z_i))}{f_Z(Z_i)}\right)\mathbf{k}_h(Z_i - Z_j) = \frac{1}{2h}\Big(\varpi_{ij} - \varpi_{ji}\Big)\mathbf{k}_h(Z_i - Z_j)$$

is the 'symmetrized' version of the first term with $\varpi_{ij} = (g(Z_i) - g(Z_j))/f_Z(Z_j)$ and ϖ_{ji} defined accordingly. Hence, the above expressions can be rewritten as symmetric second order U-statistics:

$$\begin{aligned}
&\binom{n}{2}^{-1}\sum_{i\neq j}\frac{1}{2h}(\varpi_{ij} - \varpi_{ji})\mathbf{k}_h(Z_i - Z_j) + \binom{n}{2}^{-1}\sum_{i\neq j}\frac{1}{2h}\left(\frac{V_i}{f_Z(Z_j)} - \frac{V_j}{f_Z(Z_i)}\right)\mathbf{k}_h(Z_i - Z_j) \\
=&T_{211n} + T_{212n}
\end{aligned}$$

By symmetry of the kernel function and the i.i.d. assumption on one hand, and by independence between V_i and Z_i on the other, one can straightforwardly verify that $\mathbb{E}\Big[T_{211n}\Big] = \mathbb{E}\Big[T_{212n}\Big] = 0$. Moreover, letting $r_n(Z_i, Z_j) = \frac{1}{2h}(\varpi_{ij} + \varpi_{ji})\mathbf{k}_h(Z_i - Z_j)$, one can use B2, B3, B8, and a change of variables to verify that:

$$\mathbb{E}\Big[\Big|r_n(\xi_i,\xi_j)\Big|^2\Big] = o(n)$$

Thus, since by change of variables $\mathbb{E}\Big[r_n(Z_i,Z_j)\Big|Z_i\Big] = \mathbb{E}\Big[r_n(Z_i,Z_j)\Big|Z_j\Big] = O(h)$, one can use Lemma 3.1 in Powell, Stock, and Stoker (1989) to infer that $T_{211n} = o_p\Big(\frac{1}{\sqrt{n}}\Big)$. Next, we examine the leading term T_{212n}. Let $p_n(\xi_i,\xi_j) = \frac{1}{2h}((V_i/f_Z(Z_j)) - (V_j/f_Z(Z_i)))\mathbf{k}_h(Z_i - Z_j)$ with $\xi_i = \{Z_i, V_i\}$ and $\xi_j = \{Z_j, V_j\}$. By B2, B3, B8,

and change of variables one can verify that:

$$\mathbb{E}\Big[\Big|p_n(\xi_i,\xi_j)\Big|^2\Big] = o(n)$$

Using again Lemma 3.1 in Powell, Stock, and Stoker (1989), we have that:

$$\sqrt{n}T_{212n} = \sqrt{n}\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\Big[p_n(\xi_i,\xi_j)\Big|\xi_i\Big] + o_p(1)$$

After change of variables with $u_3 = (Z_i - Z_j)/h$, independence between Z_i and V_i, and $\mathbb{E}\Big[V_j\Big|V_i\Big] = \mathbb{E}\Big[V_j\Big] = 0$:

$$\mathbb{E}\Big[p_n(\xi_i,\xi_j)\Big|\xi_i\Big] = \int\frac{1}{2}\left(\frac{V_i}{f_Z(Z_i+hu_3)} - \frac{V_j}{f_Z(Z_i)}\right)f_Z(Z_i+hu_3)f_V(V_j)dv_j = \frac{1}{2}V_i = \mathbb{E}\Big[p(\xi_i,\xi_j)\Big|\xi_i\Big]$$

where $\mathbb{E}\Big[p(\xi_i,\xi_j)\Big|\xi_i\Big]$ denotes the limit expression. Thus, we have:

$$\sqrt{n}T_{212n} = \sqrt{n}\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\Big[p(\xi_i,\xi_j)\Big|\xi_i\Big] + o_p(1) \tag{A-15}$$

Applying Lindberg Levy's Central Limit Theorem (CLT) to equation (A-15), we have:

$$\sqrt{n}T_{212n} \xrightarrow{d} N\Big(0, \int V_i^2 dF_V(V_i)\Big)$$

where $F_V(\cdot)$ is the distribution function of V_i. Thus, for (A-14) we obtain (adding the neglected term $F_V^{(1)}\Big(\overline{V}_a\Big)$):

$$\sqrt{n}\frac{1}{n}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m) - g(Z_m)\Big)\Big(F_V^{(1)}\Big(\overline{V}_a\Big) + F_V^{(1)}\Big(\overline{V}_b\Big)\Big)\int U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X}(U_G,X_k)\Big) \xrightarrow{d} N\Big(0,\Omega_1\Big)$$

where Ω_1 was defined in the statement of the lemma.

It remains to show that T_{22n} is of smaller order than the previous term. Using B2, B3, B7, a mean values expansion and a similar decomposition as for T_{12n}, one can show that $\mathbb{E}\Big[\Big|T_{22n}\Big|^\kappa\Big] = O(\|\theta-\theta_0\|h^{-(\kappa d-d)}) + o(\|\theta-\theta_0\|h^{-(\kappa d-d)})$. Thus, application of Rosenthal's inequality (with $\kappa = 1$), followed by Markov's inequality, and the bandwidth conditions imply that $T_{22n} = o_p(n^{-1})$. ■

Lemma B5. Under assumptions A1-A4, B1-B8, the term S_{222n} defined in the proof of Theorem 3 is equal to:

$$\begin{aligned}
&(\theta-\theta_0)'\frac{1}{\sqrt{n}}W_{2n} + o_p\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right) + o_p\left(\frac{1}{n}\right)\\
&= (\theta-\theta_0)'\frac{1}{n}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m)-g(Z_m)\Big)\Big(-\int I_i U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_G,X_k,V_i)\Big) + o_p\left(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\right) + o_p\left(\frac{1}{n}\right)
\end{aligned}$$

where

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\Big(\widehat{g}(Z_m) - g(Z_m)\Big)\Big(-\int I_i U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_G,X_k,V_i)\Big) \xrightarrow{d} N\Big(0,\Omega_2\Big)$$

with $\Omega_2 = E_2\Phi_2E_2'$ where

$$\Phi_2 = \Phi_1$$

and

$$E_2 = -\int I_i U_{jG}\nabla_\theta\psi_2(X_k,\theta_0)dF_{U_G,X,V}(U_{jG},X_k,V_i)$$

.

Proof of Lemma B5

Next, we consider S_{222n}. First we replace again $\widehat{f}_{X,V}(X_k,V_i)$ by $f_{X,V}(X_k,V_i)$ using B2, B3, and B8. After a mean value expansion around (V_i-V_j), we have:

$$S_{222n} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} \frac{1}{h^{d+1}} f_{X,V}^{-1}(X_k,V_i) I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(X_k,\overline{V}) \Big(\Big(\widehat{V}_i - V_i\Big) - \Big(\widehat{V}_j - V_j\Big)\Big) A_{lk}(\theta)$$

where $\mathbf{K}_{h,j}^{(1)}(x,\overline{V}_i)$ is defined in the proof of Theorem 2. As before $\Big(\widehat{V}_i - V_i\Big) = \Big(\widehat{g}(Z_i) - g(Z_i)\Big)$, while the term involving subscript j follows by an identical argument. Let $\mathbf{C}_{ijkl}(\theta) = f_{X,V}^{-1}(X_k,V_i) I_i U_{jG} \mathbf{K}_{h,j}^{(1)}(X_k,\overline{V}) \mathbf{A}_{lk}(\theta)$. Then, we have:

$$\begin{aligned} S_{212n} =& \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} \Big(\widehat{g}(Z_i) - g(Z_i)\Big) \Big(\frac{1}{h^{d+1}} \mathbf{C}_{ijkl}(\theta) - \mathbb{E}\Big[\frac{1}{h^{d+1}} \mathbf{C}_{ijkl}(\theta)\Big]\Big) \\ &+ \frac{1}{n(n-1)(n-2)(n-3)} \sum_{k\neq l\neq i\neq j} \Big(\widehat{g}(Z_i) - g(Z_i)\Big) \mathbb{E}\Big[\frac{1}{h^{d+1}} \mathbf{C}_{ijkl}(\theta)\Big] \\ =& R_{1n}(\theta) + R_{2n}(\theta) \end{aligned}$$

We start with $R_{1n}(\theta)$. Integration by parts and a similar line of argument to before can be used to show that $\mathbb{E}\Big[\Big|R_{1n}(\theta)\Big|\Big] = O\Big(\|\theta-\theta_0\| h^{-(d\kappa-d)-(\kappa-1)}\Big) + o\Big(\|\theta-\theta_0\| h^{-(d\kappa-d)-(\kappa-1)}\Big)$, while the leading term of $\mathbb{E}\Big[R_{1n}(\theta)^2\Big]$ is $O\Big(\|\theta-\theta_0\| h^{-(d+1)}\Big)$. Applying again Rosenthal's inequality yields:

$$\mathbb{E}\Big[\Big(R_{1n}(\theta)\Big)^{2\kappa}\Big] \leq n^{-8\kappa} \Xi_\kappa \Big(\|\theta-\theta_0\| \Big(n^{4\kappa} h^{-d\kappa-\kappa} + n^4 h^{-2d\kappa-2\kappa+(d+1)}\Big)\Big)$$

For $\kappa = 1$, we have $O(\|\theta-\theta_0\| n^{-4} h^{-d-1})$. By Markov's inequality and the bandwidth conditions, we have that $R_{1n}(\theta) = o_p(n^{-1})$. Next, consider $R_{2n}(\theta)$. This term only depends on i and can be shown to converge in distribution as claimed in the above lemma using the same arguments as for $T_{21n}(\theta)$ in Lemma B2. That is:

$$\sqrt{n}\frac{1}{n} \sum_{m=1}^{n} \Big(\widehat{g}(Z_m) - g(Z_m)\Big) \Big(- \int I_i U_{jG} \nabla_\theta \psi_2(X_k,\theta_0) dF_{U_G,X,V}(U_G,X_k,V_i)\Big) \xrightarrow{d} N\Big(0,\Omega_2\Big)$$

where Ω_2 was defined in the lemma. ■

Lemma B6. Under assumptions A1-A4, B1-B8, the term S_{223n} defined in the proof of Theorem 3 is equal to:

$$\begin{aligned} &(\theta-\theta_0)' \frac{1}{\sqrt{n}} W_{3n} + o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big) + o_p\Big(\frac{1}{n}\Big) \\ &\quad = (\theta-\theta_0)' \frac{1}{n} \sum_{m=1}^{n} \Big(U_{m\widehat{G}} - U_{mG}\Big) \Big(\int I_i \nabla_\theta \psi_2(X_k,\theta_0) dF_{X,V}(X_k,V_i)\Big) + o_p\Big(\frac{\|\theta-\theta_0\|}{\sqrt{n}}\Big) + o_p\Big(\frac{1}{n}\Big) \end{aligned}$$

where

$$\frac{1}{\sqrt{n}} \sum_{m=1}^{n} \Big(U_{m\widehat{G}} - U_{mG}\Big) \Big(\int I_i \nabla_\theta \psi_2(X_k,\theta_0) dF_{X,V}(X_k,V_i)\Big) \xrightarrow{d} N\Big(0,\Omega_3\Big)$$

with $\Omega_3 = E_3 \Phi_3 E_3'$ where

$$\Phi_3 = \int_0^{\phi_Y} \mathbb{E}\Big[U_{1G} I[s<U_1]\Big] H_{t1}(s) \frac{dG(s)}{(1-G(s-))}$$

and

$$E_3 = \int I_i \nabla_\theta \psi_2(X_k,\theta_0) dF_{X,V}(X_k,V_i)$$

where $F_{X,V}(\cdot,\cdot)$ denotes the joint distribution function of X_k and V_i.

Proof of Lemma B6

$\widehat{f}_{X,V}(X_k, V_i)$ in the denominator is again tackled using B2, B3, and B8. Let $\mathbf{D}_{ijkl}(\theta) = f^{-1}_{X,V}(X_k, V_i) I_i \mathbf{K}_{h,j}(X_k, V_i) A_{lk}(\theta)$. Then, S_{213n} can be decomposed as:

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\Big(U_{j\widehat{G}} - U_{jG}\Big)\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big] \\
&+ \frac{1}{n(n-1)(n-2)(n-3)}\sum_{k\neq l\neq i\neq j}\Big(U_{j\widehat{G}} - U_{jG}\Big)\Big(\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta) - \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Big) \\
=&U_{11n}(\theta) + U_{12n}(\theta)
\end{aligned}$$

Consider $U_{11n}(\theta)$. We define the following notation, which we keep as close as possible to Lu and Burke (2005):

$$\begin{aligned}
\Lambda^G(t) =& \int_{-\infty}^{t} \frac{1}{1-G(s-)} dG(s) \\
N_j(t) =& I[U_j \leq t, \Delta_j = 0] \\
M_j(t) =& N_j(t) - \int_0^t I[U_j \geq s] d\Lambda_j(s), \qquad \Lambda_j(s) = \Lambda^G(s) \\
Y_n(t) =& \sum_{j=1}^{n} I[U_j \geq t], \qquad \overline{Y}_n(t) = \frac{1}{n} Y_n(t)
\end{aligned}$$

Moreover, $F_Y(\cdot-)$ will in the following refer to the left-continuous distribution function of Y_j. Noting that U_j only has support on the positive real line and that:

$$\frac{\widehat{G}(U_j-) - G(U_j-)}{1-G(U_j-)} = \int_{s<U_j} \frac{1-\widehat{G}(s-)}{1-G(s-)} \frac{\sum_{j=1}^n dM_j(s)}{Y_n(s)} = \frac{1}{n}\int_{s<U_j} \frac{1-\widehat{G}(s-)}{1-G(s-)} \frac{\sum_{j=1}^n dM_j(s)}{\overline{Y}_n(s)}$$

$U_{11n}(\theta)$ can be rewritten as:

$$\begin{aligned}
U_{11n}(\theta) =& \frac{1}{n}\sum_{j=1}^{n}\Big(U_{j\widehat{G}} - U_{jG}\Big)\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big] \\
=& \frac{1}{n}\sum_{j=1}^{n} U_{j\widehat{G}} \frac{\widehat{G}(U_j-) - G(U_j-)}{1-G(U_j-)}\,\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big] \\
=& \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Bigg\{\frac{1}{n^2}\sum_{m=1}^{n}\int_0^\infty \sum_{j=1}^{n} \frac{U_j\Delta_j}{1-G(U_j-)} I[s<U_j] \frac{1}{\overline{Y}_n(s)} \frac{1-\widehat{G}(s-)}{1-G(s-)} dM_m(s) \\
&+ \frac{1}{n^2}\sum_{m=1}^{n}\int_0^\infty \sum_{j=1}^{n} U_j\Delta_j \left(\frac{1}{1-\widehat{G}(U_j-)} - \frac{1}{1-G(U_j-)}\right) I[s<U_j] \frac{1}{\overline{Y}_n(s)} \frac{1-\widehat{G}(s-)}{1-G(s-)} dM_m(s)\Bigg\} \\
=& U_{111n}(\theta) + U_{112n}(\theta)
\end{aligned}$$

Using Lemma A.2 (ii) in Lopez (2009) and B1, $U_{112n}(\theta)$ is of smaller order than $U_{111n}(\theta)$ and can hence be neglected in the following. Letting $H_{nt}(s) = \frac{1}{n}\sum_{j=1}^{n} U_{jG} I[s<U_j] \frac{1}{\overline{Y}_n(s)} \frac{1-\widehat{G}(s-)}{1-G(s-)}$, we have for the first term:

$$U_{111n}(\theta) = \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\frac{1}{n}\sum_{m=1}^{n}\int_0^\infty H_{nt}(s) dM_m(s)$$

Now for $0 < \nu < \phi_Y$, let:

$$U^{\nu}_{111n}(\theta) = \mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\frac{1}{n}\sum_{m=1}^{n}\int_0^{\nu} H_{nt}(s)dM_m(s)$$

Then, uniformly for $s \in [0, \nu]$, we have:

$$\begin{aligned}
H_{nt}(s) =& \frac{1}{n}\sum_{m=1}^{n} U_{jG} I[s < U_j]\frac{1}{\overline{Y}_n(s)}\frac{1-\widehat{G}(s-)}{1-G(s-)} \\
=& \frac{1}{n}\sum_{m=1}^{n} U_{jG} I[s < U_j]\frac{1}{(1-F_Y(s-))(1-G(s-))} + o_p(1) \\
=& \mathbb{E}\Big[U_{1G} I[s < U_1]\Big]\frac{1}{(1-F_Y(s-))(1-G(s-))} + o_p(1) \\
=& H_{1t}(s) + o_p(1)
\end{aligned}$$

where the second and third equality follow by adding and subtracting $\frac{1}{(1-F_Y(s-))(1-G(s-))}$ and $\mathbb{E}\Big[U_{1G} I[s < U_1]\Big]$, respectively (see Lemma A.8 in Lu and Burke (2005) for details). Moreover, using the same lines of arguments as in the proof of statement (2.29) in Lai, Ying, and Zheng (1995, p.274) and B1, we have:

$$\sqrt{n}\frac{1}{n}\sum_{m=1}^{n}\int_{\nu}^{\infty} H_{nt}(s)dM_m(s) \xrightarrow{p} 0$$

as $\nu \longrightarrow \phi_Y$ and $n \longrightarrow \infty$. Therefore:

$$\sqrt{n}U_{11n} = M_{n2t} + o_p(1)$$

For $0 < \nu < \phi_Y$, $\{M_{n2t}\}$ is a local martingale with predictable variation process (Lu and Burke, 2005, p.198).

$$\begin{aligned}
\langle M_{2nt}(\nu)\rangle =& \frac{1}{n}\sum_{m=1}^{n}\int_0^{\nu} H^2_{t1}(s) I[U_m \geq s](1-\Delta\Lambda^G(s))d\Lambda^G(s) \\
\xrightarrow{p}& \int_0^{\nu} H^2_{t1}(s)\,\mathbb{P}[U_1 \geq s](1-\Delta\Lambda^G(s))d\Lambda^G(s) \\
=& \int_0^{\nu} H^2_{t1}(s)(1-G(s-))(1-F_Y(s-))\frac{(1-G(s-)dG(s)}{(1-G(s-))(1-G(s-))} \\
=& \int_0^{\nu} \mathbb{E}\Big[U_{1G} I[s < U_1]\Big] H_{t1}(s)\frac{dG(s)}{(1-G(s-))}
\end{aligned}$$

where the second equality follows because of $\mathbb{P}[U_1 \geq s] = (1-H(s-)) = (1-G(s-))(1-F_Y(s-))$ and the definition of $\Lambda^G(s)$ before. In addition, we have:

$$\frac{1}{\sqrt{n}}\sum_{m=1}^{n}\int_{\nu}^{\phi_Y} H_{t1}(s)dM_m(s) \xrightarrow{p} 0$$

as $\nu \longrightarrow \phi_Y$. By Rebelledo's martingale central limit theorem (CLT), we obtain:

$$M_{2nt} \xrightarrow{p} N\Big(0, \Phi_3\Big)$$

with

$$\Phi_3 = \int_0^{\phi_Y} \mathbb{E}\Big[U_{1G} I[s < U_1]\Big] H_{t1}(s)\frac{dG(s)}{(1-G(s-))}$$

Thus:

$$\sqrt{n}\frac{1}{n}\sum_{m=1}^{n}\Big(U_{m\widehat{G}}-U_{mG}\Big)\Big(\int I_i\nabla_\theta\psi_2(X_k,\theta_0)dF_{X,V}(X_k,V_i)\Big)\xrightarrow{d}N\Big(0,\Omega_3\Big)$$

where Ω_3 was defined in the statement of the lemma.

It remains to show that $U_{12n}(\theta)$ is of smaller order. Notice that by a similar argument to before and B1, one can show that uniformly for $0<\nu<\phi_Y$:

$$\mathbb{E}\Big[\Big(\Big(U_{j\widehat{G}}-U_{jG}\Big)\Big(\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)-\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Big)\Big)^2\Big]=O\Big(\frac{\|\theta-\theta_0\|}{h^d}\Big)+O\Big(\|\theta-\theta_0\|^2\Big)+o(1)$$

which is $O(\|\theta-\theta_0\|h^{-d})$ by B8. A similar line of argument and B1 can used to show that the leading term of

$$\mathbb{E}\Big[\Big|\Big(U_{j\widehat{G}}-U_{jG}\Big)\Big(\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)-\mathbb{E}\Big[\frac{1}{h^d}\mathbf{D}_{ijkl}(\theta)\Big]\Big)\Big|^\kappa\Big]$$

is $O(\|\theta-\theta_0\|h^{-d(\kappa-1)})$ given B8. Thus, applying Rosenthal's inequality we obtain:

$$\begin{aligned}\mathbb{E}\Big[\Big(U_{12n}(\theta)\Big)^{2\kappa}\Big]\leq& n^{-8\kappa}\Xi_\kappa(\|\theta-\theta_0\|)\Big(n^{4\kappa}h^{-d\kappa}+n^4h^{-d(2\kappa-1)}\Big)\Big)\\=&O\Big(\|\theta-\theta_0\|n^{-4\kappa}h^{-d\kappa}\Big)+O\Big(\|\theta-\theta_0\|n^{-8\kappa+4}h^{-d(2\kappa-1)}\Big)\end{aligned}$$

By Markov's inequality, we have that $U_{12n}(\theta)=o_p(n^{-1})$ for $\kappa=1$. ■

Proof of Theorem 4

We denote by $\mathbb{E}_*$ and var_* the mean and variance operators of the bootstrapping sampling. In addition, let $O_p^*(1)$ and $o_p^*(1)$ be the orders of magnitude according to the bootstrapping distribution.

Using a similar argument to Goncalves and White (2005), the theorem follows once we show that:

$$\mathbb{E}_*\Big[\sqrt{m}(\theta^*-\widehat{\theta})\Big]=o_p(1) \tag{A-16}$$

$$\text{var}_*\Big(\sqrt{m}(\theta^*-\widehat{\theta})\Big)=\text{var}\Big(\sqrt{n}(\widehat{\theta}-\theta_0)\Big)+O_p\Big(\frac{1}{\sqrt{n}}\Big) \tag{A-17}$$

and for $\epsilon>0$:

$$\mathbb{E}_*\Big[\Big(\sqrt{m}\|\theta^*-\theta_0\|\Big)^{2+\epsilon}\Big]=O_p(1) \tag{A-18}$$

Equations (A-16) and (A-17) follow automatically once we have verified that $\sqrt{m}(\theta^*-\widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta}-\theta_0)$ up to an error of smaller order. Thus, we show that Lemma B1 and B2 are also applicable to the bootstrap estimator in (13) (with n being replaced by m in both lemmata). Since the proof is rather lengthy and in large parts identical to before, we will only sketch the one of asymptotic normality and $\sqrt{m}$-consistency paralleling the proof of Theorem 3. Consistency follows in fact by similar arguments to the proof of Theorem 2 and the ones presented in the following.

The equation in (13) can be decomposed as in the proof of Theorem 3. That is, we examine:

$$\frac{1}{m(m-1)}\sum_{k\neq l}\mu(X_k^*)\times A_{lk}^*(\theta)+\frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widehat{\mu}^*(X_k^*)-\mu(X_k^*)\Big)\times A_{lk}^*(\theta)$$

with $A_{lk}^*(\theta)=I[X_k^*\in\mathcal{X}]\big\{I[X_k^{*\prime}\beta(\theta)>X_l^{*\prime}\beta(\theta)]-I[X_k^{*\prime}\beta(\widehat{\theta})>X_l^{*\prime}\beta(\widehat{\theta})]\big\}$. In a first step we show that:

$$\frac{1}{m(m-1)}\sum_{k\neq l}\mu(X_k^*)A_{lk}^*(\theta) \tag{A-19}$$

behaves as

$$S_{1n} = \frac{1}{n(n-1)} \sum_{k \neq l} \mu(X_k) A_{lk}(\theta)$$

from the proof of Theorem 3. Since (A-19) is again a second order U-statistic for every $\theta \in \Theta$, the same Hoeffding decomposition argument as in Lemma A.3 of Khan (2001) used in the proof of Theorem 3 can be applied: first notice that the conditional expectation over bootstrap samples given X_k and X_l, respectively, is:

$$\begin{aligned} \psi^*(X_k^*, \theta) =& \frac{1}{2}\Big\{ \mathbb{E}_*\Big[\mu(X_k^*) A_{lk}^*(\theta)\Big| X_k^*\Big] + \mathbb{E}_*\Big[\mu(X_l^*) A_{kl}^*(\theta)\Big| X_k^*\Big]\Big\} \\ =& \frac{1}{2}\Big\{ \frac{1}{n}\sum_{l=1}^{n} \mu(X_k^*) A_{lk^*}(\theta) + \frac{1}{n}\sum_{l=1}^{n} \mu(X_l) A_{lk^*}(\theta)\Big\} \end{aligned} \tag{A-20}$$

where the subscript without star in the second line indicates the summable variable. Hence:

$$\mathbb{E}_*\Big[\psi^*(X_k^*, \theta)\Big] = \frac{1}{2}\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\Big\{\mu(X_k) A_{lk}(\theta) + \mu(X_l) A_{lk}(\theta)\Big\}$$

This term can be expanded further to give:

$$\mathbb{E}\Big[\psi(X_k, \theta)\Big] + \Big(\frac{1}{n}\sum_{k=1}^{n} \psi^*(X_k, \theta) - \mathbb{E}\Big[\psi(X_k, \theta)\Big]\Big) = T_{1n}^* + T_{2n}^* \tag{A-21}$$

T_{1n}^* can be expanded as in Lemma A.3 of Khan (2001). For T_{2n}^* on the other hand, let $\phi_{lk}(\theta) = \Big\{\mu(X_k) A_{lk}(\theta) + \mu(X_l) A_{lk}(\theta)\Big\} - \mathbb{E}\Big[\Big\{\mu(X_k) A_{lk}(\theta) + \mu(X_l) A_{lk}(\theta)\Big\}\Big]$. Notice that by B2, B5, and B6 we have $\mathbb{E}\Big[\Big|\phi_{lk}(\theta)\Big|^{\kappa}\Big] = \|\theta - \theta_0\|$ and $\mathbb{E}\Big[\phi_{lk}(\theta)^2\Big] = \|\theta - \theta_0\|$. Thus, by Rosenthal's inequality:

$$\mathbb{E}\Big[{T_{2n}^*}^{2\kappa}\Big] \leq n^{-4\kappa}\Xi_{\kappa}\Big(\|\theta - \theta_0\| n^{2\kappa} + \|\theta - \theta_0\| n^2\Big)$$

For $\kappa = 1$, $\mathbb{E}\Big[{T_{2n}^*}^{2\kappa}\Big] = O(\|\theta - \theta_0\| n^{-2})$ and thus by Markov's inequality $T_{2n}(\theta) = o_p\Big(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\Big)$. Next, we show that 'm out of n' bootstrap is also able to mimic the random elements of the 'projection' of the U-statistic used in Lemma A.3 of Khan (2001). Notice that:

$$\begin{aligned} &\frac{1}{m}\sum_{i=1}^{m}\Big\{\psi^*(X_k^* = X_i^*, \theta) - \mathbb{E}_*\Big[\psi^*(X_k^* = X_i^*, \theta)\Big]\Big\} \\ =& \frac{1}{m}\sum_{i=1}^{m}\Big\{\psi(X_k^* = X_i^*, \theta) - \mathbb{E}\Big[\psi(X_k^* = X_i^*, \theta)\Big]\Big\} + o_p(1) \end{aligned}$$

where the $o_p(1)$ term follows again by a subsequent application of Rosenthal's and Markov's inequality. The term in curley brackets can be dealt with by the same arguments as in Lemma A.3 of Khan (2001) since $m \longrightarrow \infty$ as $n \longrightarrow \infty$. Finally, the quadratic term of the Hoeffding decomposition can be shown to be $o_p\Big(\frac{1}{m}\Big)$ by expanding this term as above and subsequently applying Rosenthal's and Markov's inequality.

Next, we give a rough sketch of the steps to show that:

$$\frac{1}{m(m-1)}\sum_{k \neq l}\Big(\widehat{\mu}^*(X_k^*) - \mu(X_k^*)\Big) A_{lk}^*(\theta) \tag{A-22}$$

behaves as

$$S_{2n} = \frac{1}{n(n-1)}\sum_{k \neq l}\Big(\widehat{\mu}(X_k) - \mu(X_k)\Big) A_{lk}(\theta)$$

A similar decomposition as in the proof of Theorem 3 yields:

$$\begin{aligned}&\frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widetilde{\mu}^*(X_k^*)-\mu(X_k^*)\Big)A_{lk}^*(\theta)+\frac{1}{m(m-1)}\sum_{k\neq l}\Big(\widehat{\mu}^*(X_k^*)-\widetilde{\mu}^*(X_k^*)\Big)A_{lk}^*(\theta)\\ =&S_{21n}^*+S_{22n}^*\end{aligned}$$

where $\widetilde{\mu}^*(\cdot)$ is defined analogously to the proof of Theorem 3. We start with S_{21n}^*, which can again be rewritten as:

$$S_{21n}^*=\frac{1}{m(m-1)}\sum_{k\neq l}\left\{\frac{1}{m}\sum_{i=1}^m\frac{\frac{1}{mh^{*d}}\sum\limits_{j=1}^m I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j}(X_k^*,V_i^*)}{\frac{1}{mh^{*d}}\sum\limits_{j=1}^m I_i^*\mathbf{K}_{h^*,j}(X_k^*,V_i^*)}\right\}\times A_{lk}^*(\theta)$$

Using B2, B3, B8, and the same argument as in the proof of Lemma B3, we can replace $\widehat{f}_{X,V}^*(X_k^*,V_i^*)$ by $f_{X,V}(X_k^*,V_i^*)$. After omitting terms with $k=l=j=i$, which, parallel to Theorem 3, results in an error of order $o_p^*(\|\theta-\theta_0\|/mh^{*d})$), the numerator is given by:

$$\frac{1}{m(m-1)(m-2)(m-3)}\sum_{k\neq l\neq i\neq j}\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\tag{A-23}$$

Using a similar decomposition as in (A-21), it is straightforward to show that this fourth order U-statistic is degenerate in ξ_k^*,ξ_l^*,ξ_i^* for each $\theta\in\Theta$, where ξ_k^*,ξ_l^*,ξ_i^* are defined as in the proof of Lemma B3. As before, this also implies that:

$$\begin{aligned}&\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big]\\ =&\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]\\ &+\left\{\frac{1}{n^4h^{*d}}\sum_{k=1}^n\sum_{l=1}^n\sum_{j=1}^n\sum_{i=1}^n\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h^*,j}(X_k,V_i)A_{lk}(\theta)\right.\\ &\left.-\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]\right\}\\ =&o_p\Big(\frac{1}{n}\Big)\end{aligned}$$

where the last equality follows by B8, the Rosenthal's and Markov's inequalities, and the fact that $\mathbb{E}\Big[\frac{1}{h^d}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{jG}-\mu(X_k)\big)\mathbf{K}_{h,j}(X_k,V_i)A_{lk}(\theta)\Big]=0$. The second term of the Hoeffding projection of (A-23) yields:

$$\begin{aligned}&\frac{1}{m}\sum_{p=1}^m\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big|\xi_j^*=\xi_p^*\Big]\\ =&\frac{1}{m}\sum_{p=1}^m\int\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{pG}^*-\mu(X_k)\big)\mathbf{K}_{h^*,p^*}(X_k,V_i)A_{lk}(\theta)dF_{X,V}(X_k,X_l,V_i)+o_p\Big(\frac{1}{m}\Big)\end{aligned}$$

where $o_p\Big(\frac{1}{m}\Big)$ follows by another application of Rosenthal's inequality. The term $\int\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k,V_i)}I_i\big(U_{pG}^*-\mu(X_k)\big)\mathbf{K}_{h^*,p^*}(X_k,V_i)A_{lk}(\theta)dF_{X,V}(X_k,X_l,V_i)$ can now be treated as in the proof of Theorem 3 using a second order Taylor expansion:

$$(\theta-\widehat{\theta})'I_i\big(U_{iG}^*-\mu(X_k)\big)\nabla_\theta\psi_2(X_k,\widehat{\theta})+O(\|\theta-\widehat{\theta}\|^2)$$

In view that:

$$\begin{aligned}
&\operatorname{var}_*\Big(\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big|\xi_j^*\Big]\Big)\\
=&\operatorname{var}_*\Big(\mathbb{E}_*\Big[\frac{1}{h^{*d}}\frac{1}{f_{X,V}(X_k^*,V_i^*)}I_i^*\big(U_{jG}^*-\mu(X_k^*)\big)\mathbf{K}_{h^*,j^*}(X_k^*,V_i^*)A_{lk}^*(\theta)\Big|\xi_j^*\Big]\Big)\\
=&\Omega_0+o_p(1)
\end{aligned}$$

and since all higher order degenerate U-statistics from the decomposition are of smaller order, the term in (A-23) weakly converges to $N(0,\Omega_0)$ as both m and n go to infinity thus mimicking the limiting distribution of (A-13). S_{22n}^* and the three leading terms arising in analogy to S_{22n} can be treated as in the proof of Lemmata B4 to B6 using similar arguments to above. That is, the same decomposition as in the proof of those lemmata yields the remaining variance pieces Ω_1, Ω_2, and Ω_3 as m and n go to infinity. It follows that $\sqrt{m}(\theta^*-\widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta}-\theta_0)$.

To verify equation (A-18), notice that from Lemma B1 and B2, $(\theta^*-\widehat{\theta})$ can be restated as:

$$(\theta^*-\widehat{\theta})=-J^{-1}\frac{1}{\sqrt{m}}W_m+o_p\Big(\frac{1}{m}\Big)$$

where the $o_p(m^{-1})$ follows from $\sqrt{m}$-consistency. Thus, $\sqrt{m}(\theta^*-\widehat{\theta})=-J^{-1}W_m+o_p\Big(\frac{1}{\sqrt{m}}\Big)$. Recalling that $J=\frac{1}{2}\,\mathbb{E}\big[\nabla_{\theta\theta'}\psi_1(X_k,\widehat{\theta})\big]$ is bounded and negative definite by B6, we can bound equation (A-18) as follows:

$$\begin{aligned}
\mathbb{E}_*\Big[\Big(\sqrt{m}\|\theta^*-\theta_0\|\Big)^{2+\epsilon}\Big]=&\mathbb{E}_*\Big[\Big(\|-J^{-1}W_m\|\Big)^{2+\epsilon}\Big)\Big]+o_p(1)\\
\leq&\Xi_J\|W_n\|^{2+\epsilon}=O_p(1)
\end{aligned}$$

where Ξ_J is a generic constant and the last equality follows since W_n converges in distribution.

■

Appendix C

Table 1: Monte Carlo Simulation - No Censoring

Design I					
No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
$n = 100$	RankCF	0.1595	0.1400	0.2853	0.2246
	MRE	0.6382	0.6000	0.6823	0.6382
	MRC	−0.5691	−0.6000	1.6359	1.4424
$n = 200$	RankCF	0.1173	0.1100	0.2279	0.1828
	MRE	0.6155	0.6000	0.6346	0.6155
	MRC	−0.5591	−0.6300	1.6651	1.4857
$n = 400$	RankCF	0.0969	0.0800	0.1974	0.1576
	MRC	0.6078	0.6000	0.6167	0.6078
	MRE	−0.5617	−0.6100	1.7220	1.5406
$n = 600$	RankCF	0.0831	0.0800	0.1865	0.1487
	MRC	0.6024	0.6000	0.6083	0.6024
	MRE	−0.5378	−0.5400	1.7330	1.5538
Design II					
$n = 100$	RankCF	0.2501	0.1300	0.5965	0.4332
	MRE	0.6482	0.5800	0.8200	0.6709
	MRC	0.6324	0.5400	0.8181	0.6684
$n = 200$	RankCF	0.1438	0.1000	0.3852	0.2743
	MRE	0.6402	0.5900	0.7431	0.6423
	MRC	0.6434	0.6100	0.7505	0.6480
$n = 400$	RankCF	0.1053	0.0800	0.3040	0.2219
	MRE	0.6326	0.5900	0.6995	0.6326
	MRC	0.6335	0.5700	0.7042	0.6338
$n = 600$	RankCF	0.1254	0.1000	0.2833	0.2128
	MRE	0.6205	0.5900	0.6644	0.6205
	MRC	0.6244	0.5900	0.6710	0.6244

[1] The figures in the table represent the average of the corresponding bias measure (401 replications).

Table 2: Monte Carlo Simulation - Censoring

Design III					
No. of obs.	Estimator	Mean Bias[1]	Median Bias[1]	RMSE[1]	MAD[1]
	RankCF	0.2260	0.1000	0.8289	0.6105
$n = 100$	MRE	0.6185	0.5900	0.8825	0.7236
(Avg. Censoring Ratio: .16)	MRC	0.6759	0.6300	0.7776	0.6801
	RankCF	0.2343	0.1000	0.6207	0.4461
$n = 200$	MRE	0.6452	0.5600	0.8329	0.6830
(Avg. Censoring Ratio: .16)	MRC	0.6519	0.5900	0.7205	0.6519
	RankCF	0.1608	0.1100	0.4568	0.3364
$n = 400$	MRC	0.6326	0.5700	0.7591	0.6377
(Avg. Censoring Ratio: .16)	MRE	0.6040	0.5800	0.6322	0.6040
	RankCF	0.1475	0.0900	0.4133	0.2948
$n = 600$	MRC	0.6611	0.6100	0.7555	0.6614
(Avg. Censoring Ratio: .16)	MRE	0.6090	0.6000	0.6272	0.6090
Design IV					
	RankCF	0.2118	0.1300	1.1672	0.9582
$n = 100$	MRE	0.4892	0.5900	1.0374	0.8670
(Avg. Censoring Ratio: .32)	MRC	0.5029	0.5800	1.0283	0.8475
	RankCF	0.1995	0.1000	0.8903	0.6741
$n = 200$	MRE	0.5129	0.5200	0.8907	0.7181
(Avg. Censoring Ratio: .32)	MRC	0.5757	0.5600	0.8495	0.6997
	RankCF	0.2318	0.1200	0.7608	0.5674
$n = 400$	MRE	0.6845	0.6200	0.8909	0.7345
(Avg. Censoring Ratio: .33)	MRC	0.6779	0.6300	0.7923	0.6870
	RankCF	0.2168	0.1000	0.6140	0.4479
$n = 600$	MRE	0.6326	0.5800	0.8109	0.6566
(Avg. Censoring Ratio: .32)	MRC	0.6206	0.5900	0.7059	0.6217

[1] The figures in the table represent the average of the corresponding bias measure (401 replications).

Table 3: Empirical Illustration - Earnings Study

Estimator	Coefficient[1]	Value	95% Bootstrap-CI
RankCF	Constant	–	
	Education	0.1121	[0.0197;0.3399]
	Age	−0.0218	[−0.0990;−0.0080]
MRE	Constant	–	
	Education	0.1949	[0.0885;0.3110]
	Age	−0.0953	[−0.1540;−0.0579]
MRC	Constant	–	
	Education	0.2162	[0.1391;0.3122]
	Age	−0.0802	[−0.1229;−0.0480]
OLS	Constant	19.8393	[14.5055;28.9085]
	Education	0.1987	[0.1046;0.3256]
	Age	−0.1003	[−0.1555;−0.0552]
LAD	Constant	19.8728	[15.1346;28.4499]
	Education	0.2229	[0.1456;0.3657]
	Age	−0.0805	[−0.1359;−0.0480]
TSLS	Constant	17.3046	[12.4894;24.9949]
	Education	0.1879	[−0.0052;0.3811]
	Age	−0.0890	[−0.1461;−0.0522]

[1] The gender coefficient has been normalized to one.

References

Abrevaya, J. (1999): "Rank Regression for Current-status data: Asymptotic normality," *Statistics and Probability Letters*, 43, 275–287.

——— (2000): "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 95(1), 1–23.

Abrevaya, J., and J. A. Hausman (1999): "Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells.," *Annales d'Economie et de Statistique*, 55/56, 243–275.

——— (2004): "Response error in a transformation model with an application to earnings-equation estimation," *Econometrics Journal*, 7, 366–388.

Arcones, M., and E. Gine (1992): "On the bootstrap of U and V statistics," *Annals of Statistics*, 20, 655–674.

Augustin, T. (1999): "Correcting for measurement error in parametric duration models by quasi-likelihood," Discussion Paper 157, Sonderforschungsbereich 386, Universität München.

Bound, B., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 59, pp. 3705–3843. Elsevier.

Bound, J., and A. Krueger (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics*, 9, 1–24.

Bricker, J., and G. V. Engelhardt (2007): "Measurement Error In Earnings Data In The Health and Retirement Study," Working Paper 16, Centre for Retirement Research at Boston College.

Butcher, K., and A. Case (1994): "The effect of sibling sex composition on women's education and earnings," *Quarterly Journal of Economics*, 109(3), 531–563.

Card, D. (1999): "The causal effect of education on earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3. Elsevier Science.

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

Cavanagh, C., and R. Sherman (1998): "Rank estimation for monotonic index models," *Journal of Econometrics*, 84, 351–381.

Chalak, K., and H. White (2007): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," Working paper, UCSD.

Chen, K., and S.-H. Lo (1997): "On the rate of uniform convergence of the product-limit estimator: strong and weak laws," *Annals of Statistics*, 25, 1050–1087.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72(2), 343–366.

Chesher, A., M. Dumangane, and R. J. Smith (2002): "Duration response measurement error," *Journal of Econometrics*, 111, 169–1994.

Corradi, V., W. Distaso, and N. Swanson (2010): "Predictive Inference for Integrated Volatility," Working paper, Warwick University.

Cristia, J., and J. Schwabish (2007): "Measurement Error in the SIPP: Evidence from matched administrative records," Working paper, Congressional Budget Office.

Dumangane, M. (2007): "Measurement Error Bias Reduction in Unemployment Durations," Discussion paper, Cemmap Working Paper.

Glewwe, P., and H. Patrinos (1999): "The Role of the Private Sector in Education in Vietnam: Evidence From the Vietnam Living Standards Survey," *World Development*, 27(5), 887–902.

Goncalves, S., and H. White (2004): "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119, 199–219.

——— (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, 100(471), 970–979.

Hahn, J., Y. Hu, and G. Ridder (2008): "Instrumental Variable Estimation of Nonlinear Models with Nonclassical Measurement Error Using Control Variates," Work in progress, University of Southern California.

Han, A. (1987): "Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator.," *Journal of Econometrics*, 35, 303–316.

Hoderlein, S., and J. Winter (2007): "Recall Errors in Surveys," Discussion paper, University of Mannheim, mimeo.

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

Ichino, A., and R. Winter-Ebmer (1999): "Lower and Upper Bounds of Returns to Schooling: An Exercice in IV Estimation with Dierent Instruments," *European Economic Review*, 43(6), 889–902.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, 77(5), 1481–1512.

Jaeckle, A. (2008): "Measurement Error and Data Collection Methods: Effects on Estimates from Event History Data," Working Paper 2008-13, Institute for Social and Economic Research (ISER).

Jochmans, K. (2010): "Estimating Monotone Index Models with Nonparametric Controls," Job-Market Paper.

Kaplan, E., and P. Meier (1958): "Nonparametric Estimation for Incomplete Observations," *Journal of American Statistical Association*, 53, 457–481.

Khan, S. (2001): "Two-stage rank estimation of quantile index models," *Journal of Econometrics*, 100(2), 319–355.

Koul, H., V. Susarla, and J. van Ryzin (1981): "Regression analysis with randomly right-censored data," *Annals of Statistics*, 9, 1276–1288.

Lai, T., Z. Ying, and Z. Zheng (1995): "Asymptotic Normality of a class of adaptive statistics with applications to synthetic data methods for censored regression," *Journal of Multivariate Analysis*, 52, 259–279.

Lopez, O. (2009): "Single-index regression models with right-censored responses," *Journal of Statitical Planning and Inference*, 139, 1082–1097.

Lu, X., and M. Burke (2005): "Censored multiple regression by the method of average derivatives," *Journal of Multivariate Analysis*, 95, 182–205.

Lu, X., and T. Cheng (2007): "Randomly censored partially linear single-index models," *Journal of Multivariate Analysis*, 98, 1895–1922.

Matzkin, R. L. (2007): "Nonparametric Survey Response Errors," *International Economic Review*, 48(4), 1411–1427.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Serfling, R. (1980): *Approimation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics. John Wiley, New York.

Sherman, R. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61(1), 123–137.

Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Skinner, C., and K. Humphreys (1999): "Weibull Regression for Lifetimes measured with Error," *Lifetime Data Analysis*, 5, 23–37.

Srinivasan, C., and M. r. w. c. Zhou (1994): "Linear regression with censoring," *Journal of Multivariate Analysis*, 49, 179–201.

Torelli, N., and U. Trivellato (1989): "Youth unemployment duration from the Italian labor force survey," *European Economic Review*, 33, 407–415.